

FACTORES DEMOGRÁFICOS, ACADÉMICOS Y SOCIOECONÓMICOS QUE INFLUYEN EN LOS RESULTADOS DEL COMPONENTE GENÉRICO DE LA PRUEBA SABER PRO

**CASO: LICENCIATURA EN MATEMÁTICAS Y ESTADÍSTICA DE LA
UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA
FACULTAD SECCIONAL DUITAMA**

Jorber Andrés Pardo Franco

Código: 201021494

Trabajo de grado presentado como requisito parcial para optar al título de:
Licenciado en Matemáticas y Estadística

Directora:

M.Sc. Carmen Helena Cepeda Araque

Universidad Pedagógica y Tecnológica de Colombia
Facultad Seccional Duitama, Licenciatura en Matemáticas y Estadística
Duitama - Boyacá, Colombia

2017

2017

Nota de aceptación

Firma jurado 1

Firma jurado 2

Firma presidente del jurado

Duitama, Octubre del 2017

*“Los grandes resultados
requieren grandes ambiciones”
Heráclito*

Agradecimientos

Mis más sinceros agradecimientos a:

A mi familia, en especial a mi Madre y Hermano, quienes me apoyaron para seguir adelante. A mis compañeros y profesores.

A mi Directora de Monografía docente de la Universidad Pedagógica y Tecnológica de Colombia, quien me orientó para llevar a cabo los propósitos de éste trabajo con la mayor disposición y paciencia, y también porque me compartió sus conocimientos estadísticos.

TÍTULO: FACTORES DEMOGRÁFICOS, ACADÉMICOS Y SOCIOECONÓMICOS QUE INFLUYEN EN LOS RESULTADOS DEL COMPONENTE GENÉRICO DE LA PRUEBA SABER PRO

CASO: LICENCIATURA EN MATEMÁTICAS Y ESTADÍSTICA DE LA UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA FACULTAD SECCIONAL DUITAMA

AUTOR: PARDO FRANCO, Jorber Andrés.

DIRECTOR (A): CEPEDA ARAQUE, Carmen Helena.

PUBLICACIÓN: Duitama, Universidad Pedagógica y Tecnológica de Colombia, 2017.

UNIDAD PATROCINANTE: Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama, Licenciatura en Matemáticas y Estadística.

PALABRAS CLAVES: Factores, académicos, demográficos, socioeconómicos, puntaje, desempeño, análisis bivariado, modelo óptimo, modelamiento estadístico, modelos aditivos generalizados para ubicación, escala y forma, regresión logística ordinal.

OBJETIVO: Determinar los factores académicos, demográficos y socioeconómicos que inciden en los resultados del componente genérico de las pruebas Saber Pro presentada por los estudiantes de Licenciatura en Matemáticas y Estadística.

DESCRIPCIÓN: En el documento se presenta un estudio que tiene como eje principal determinar los factores influyentes en el puntaje de las pruebas Saber Pro presentadas por estudiantes del programa Licenciatura en Matemáticas y Estadística desde el segundo semestre del 2011 al 2015, teniendo como fuente principal de información las bases de datos del ICFES.

FUENTES: Para el desarrollo de esta investigación se consultaron cinco libros de estadística, específicamente de análisis bivariado, modelos aditivos generalizados para ubicación, escala y forma (GAMLSS), regresión logística ordinal (RLO), además se con-

sultaron 12 artículos, en los cuales se describían estudios referentes a los factores que influyen en los resultados de las pruebas Saber Pro. Se trabajó con las bases de datos suministradas por el Instituto Colombiano para la Evaluación de la Educación (ICFES).

CONTENIDO: La investigación se describe en cinco capítulos, en el primero se presenta el contexto que orientó la investigación, en el segundo se encuentra el marco de referencia sobre pruebas Saber Pro y modelamiento estadístico, en el tercero se presenta la metodología de trabajo, el cuarto muestra el análisis de la información basados en los modelos de Regresión Logística Ordinal (RLO) y Modelos Aditivos Generalizados para Ubicación, Escala y Forma (GAMLSS) y un quinto capítulo que contiene la presentación de los resultados, finalizando con las conclusiones y discusión.

METODOLOGÍA: El estudio se clasifica como una investigación descriptiva, ya que se interpretó los datos brindados por el ICFES en el programa de Licenciatura en Matemáticas y Estadística, buscando determinar los factores demográficos, socioeconómicos y académicos que inciden en el resultado del componente genérico de las pruebas Saber Pro. El estudio presentó un enfoque cuantitativo, ya que para la interpretación se utilizaron herramientas estadísticas tales como el análisis descriptivo y bidimensional, modelamiento por medio de los GAMLSS y RLO.

CONCLUSIONES: Se encontró que la educación de los padres (primaria y secundaria) influye de manera negativa en el puntaje de cada componente, mientras que en los desempeños en Comunicación Escrita e Inglés no se presentó incidencia por estas variables.

Se observó que pertenecer al sexo masculino, ser cabeza de familia, y tener: mayor edad, una situación del hogar temporal, un número mayor de personas a cargo y en el hogar, aumenta el puntaje de Razonamiento Cuantitativo.

Para el puntaje de Inglés, se observa que ser casado, tener mayor edad y una situación del hogar temporal, aumenta el puntaje, mientras que ser hombre y tener un número mayor de personas en el hogar lo disminuye. El desempeño en Inglés decrece en todos los niveles al tener mayor edad.

Respecto a Lectura Crítica, se encontró que ser cabeza de familia y tener una mayor edad disminuye el puntaje, en cambio tener un número mayor de personas a cargo lo aumenta, como también tener televisión y un salario mensual familiar entre 4 y 5 S.M.V, aumenta el puntaje.

El ser cabeza de familia casado o soltero con mayor edad, y tener un número mayor de personas en el hogar, disminuye el puntaje de Comunicación Escrita, mientras que tener un número mayor de personas a cargo lo aumenta. Contar con servicio de internet aumenta el desempeño en Comunicación Escrita, mientras que tener teléfono fijo lo disminuye.

Se observó que las variables del Factor Socioeconómico no influyen en los puntajes de Razonamiento Cuantitativo, Comunicación Escrita y el desempeño en Inglés.

Respecto al puntaje de Inglés, se encontró que pertenecer al estrato dos, donde la ocupación de la madre es obrero, y tener: celular, un mayor número de dormitorios y horas de trabajo, lavadora, trabajo, un valor de matrícula menor a 500,000 pesos y otro de tres a cinco millones de pesos, disminuye el puntaje.

De acuerdo con Castellanos *et al* (2014), existe una diferencia entre los puntajes obtenidos en Inglés por los estudiantes de diferentes estratos, como también una asociación con el desempeño en Comunicación Escrita. Además se encontró una diferencia en los promedios, respecto al sexo, en Razonamiento Cuantitativo, obteniendo puntajes más altos las mujeres en este componente, de manera opuesta como lo menciona Castellanos *et al* (2014).

Al comparar resultados con Manrique, J. y Carreño, C. (2014), el sexo no afecta el desempeño en Comunicación Escrita, es decir no importa ser hombre o mujer para obtener niveles alto o bajos, ya que no es una variable influyente. Para el caso de internet, resulta tener una influencia positiva en el desempeño de Comunicación Escrita como concluyeron Manrique, J. y Carreño, C. (2014). De manera distinta con Manrique, J. y Carreño, C. (2014), la educación del padre (Primaria y Secundaria), si presento incidencia en el promedio del desempeño en Comunicación escrita y en el puntaje de Inglés.

Se recomienda una incorporación de Inglés en el plan de estudios de tal manera que influya en el promedio acumulado del plan de estudios, esto sería lo adecuado para fortalecer el desempeño en esta área y así lograr mejores resultados en este componente de la pruebas Saber Pro. Por otro lado, el internet es una herramienta relevante para el desarrollo general de las temáticas, contar con un apto funcionamiento de este en la Universidad, sería de gran ayuda para el desempeño de los estudiantes de Licenciatura en Matemáticas y Estadística.

Contenido

1. PRESENTACIÓN	1
2. REFERENTES TEÓRICOS	7
2.1. ANTECEDENTES	7
2.2. EL CONTEXTO DE LA EVALUACIÓN	8
2.2.1. Objetivos de los exámenes - Saber Pro	12
2.2.2. Competencias genéricas	13
2.3. MODELOS LINEALES GENERALIZADOS (GLM)	20
2.4. MODELOS ADITIVOS GENERALIZADOS PARA LOCALIZACIÓN, ESCALA Y FORMA (GAMLSS)	22
2.4.1. Algunas distribuciones asumidas en los GAMLSS	26
2.4.2. Selección del modelo	27
2.4.3. Diagnóstico del modelo	28
2.5. REGRESIÓN LOGÍSTICA ORDINAL (RLO)	30
3. DISEÑO METODOLÓGICO	34
3.1. POBLACIÓN Y TIPO DE DISEÑO	35
3.2. MÉTODOS DE ANÁLISIS Y PROCESAMIENTO DE DATOS	35
3.3. VARIABLES DE ESTUDIO	36
4. DESCRIPCIÓN DE LOS DATOS	42
4.1. RESULTADOS COMPETENCIAS GENÉRICAS	43
4.1.1. Resultados en competencias genéricas año 2016	49
4.2. Algunas características de los estudiantes que presentan la prueba . . .	50

5. FACTORES ASOCIABLES A LOS RESULTADOS DEL COMPONENTE GENÉRICO	58
5.1. ESPECIFICACIÓN DE LOS MODELOS	58
5.1.1. Puntaje Razonamiento Cuantitativo	59
5.1.2. Puntaje Inglés	61
5.1.3. Puntaje Lectura Crítica	63
5.1.4. Puntaje Comunicación Escrita	65
5.2. COMPONENTE SISTEMÁTICO	66
5.3. SELECCIÓN, EVALUACIÓN E INTERPRETACIÓN DEL MODELO ÓPTIMO	69
5.3.1. Puntaje Razonamiento Cuantitativo	70
5.3.2. Puntaje Inglés	82
5.3.3. Puntaje Lectura Crítica	97
5.3.4. Puntaje Comunicación Escrita	108
5.3.5. Desempeño Comunicación Escrita	121
5.3.6. Desempeño Inglés	123
6. DISCUSIÓN	126
7. CONCLUSIONES	130
REFERENCIAS BIBLIOGRÁFICAS	133

Lista de Figuras

2-1. Algunas funciones de enlace	22
2-2. Gráficos salida plot()	29
2-3. Parámetros salida plot()	29
3-1. Variables de estudio Factor Académico.	37
3-2. Variables de estudio Factor Académico.	38
3-3. Variables de estudio Factor Demográfico.	39
3-4. Variables de estudio Factor Socioeconómico.	40
3-5. Variables de estudio Factor Socioeconómico.	41
4-1. Desempeño Competencias Genéricas.	48
5-1. Datos atípicos en puntaje Razonamiento Cuantitativo.	59
5-2. Selección de distribución para puntaje Razonamiento Cuantitativo. . .	59
5-3. Ajuste puntaje Razonamiento Cuantitativo.	60
5-4. Datos atípicos en puntaje Inglés.	61
5-5. Selección de distribución para puntaje Inglés.	61
5-6. Ajuste puntaje de Inglés.	62
5-7. Datos atípicos en puntaje Lectura Crítica.	63
5-8. Selección de distribución para puntaje Lectura Crítica.	63
5-9. Ajuste puntaje Lectura Crítica.	64
5-10. Datos atípicos en puntaje Comunicación Escrita.	65
5-11. Selección de distribución para puntaje Comunicación Escrita.	65
5-12. Ajuste puntaje Comunicación Escrita.	66
5-13. Modelo óptimo Razonamiento Cuantitativo Factor Académico.	70
5-14. Diagnóstico de la distribución modelo Razonamiento Cuantitativo Factor Académico.	71
5-15. Prueba ajuste modelo Razonamiento Cuantitativo Factor Académico. .	71

5-16.Ajuste de la distribución modelo Razonamiento Cuantitativo Factor Académico.	72
5-17.Modelo óptimo Razonamiento Cuantitativo Factor Demográfico.	73
5-18.Diagnóstico de la distribución modelo Razonamiento Cuantitativo Factor Demográfico.	74
5-19.Prueba ajuste modelo Razonamiento Cuantitativo Factor Demográfico.	74
5-20.Ajuste de la distribución modelo Razonamiento Cuantitativo Factor Demográfico.	75
5-21.Valores Atípicos Número de personas a cargo Razonamiento Cuantitativo.	75
5-22.Valores Atípicos Edad Razonamiento Cuantitativo.	75
5-23.Valores atípicos Número personas hogar Razonamiento Cuantitativo.	76
5-24.Prueba de valores atípicos número de personas a cargo Razonamiento Cuantitativo.	76
5-25.Prueba de valores atípicos Edad Razonamiento Cuantitativo.	77
5-26.Prueba de valores atípicos Número de personas en el hogar Razonamiento Cuantitativo.	77
5-27.Modelo óptimo Razonamiento Cuantitativo Factor Socioeconómico.	80
5-28.Diagnóstico de la distribución modelo Razonamiento Cuantitativo Factor Socioeconómico.	81
5-29.Prueba ajuste modelo Razonamiento Cuantitativo Factor Socioeconómico.	81
5-30.Ajuste de la distribución modelo Razonamiento Cuantitativo Factor Socioeconómico.	82
5-31.Modelo óptimo Inglés Factor Académico.	83
5-32.Diagnóstico de la distribución modelo Inglés Factor Académico.	84
5-33.Prueba de ajuste modelo Inglés Factor Académico.	84
5-34.Ajuste de la distribución modelo Inglés Factor Académico.	85
5-35.Modelo óptimo Inglés Factor Demográfico.	85
5-36.Modelo óptimo Inglés Factor Demográfico.	86
5-37.Diagnóstico de la distribución modelo Inglés Factor Demográfico.	87
5-38.Prueba ajuste modelo Inglés Factor Demográfico.	87
5-39.Ajuste de la distribución modelo Inglés Factor Demográfico.	88
5-40.Valores atípicos Número de personas a cargo Inglés.	88
5-41.Valores atípicos Edad Inglés.	88
5-42.Valores atípicos Número personas hogar Inglés.	89
5-43.Prueba de valores atípicos número de personas a cargo Inglés.	89
5-44.Prueba de valores atípicos Edad Inglés.	90
5-45.Prueba de valores atípicos Número de personas en el hogar Inglés.	90

5-46.Modelo óptimo Inglés Factor Socioeconómico.	92
5-47.Modelo óptimo Inglés Factor Socioeconómico.	93
5-48.Diagnóstico de la distribución modelo Inglés Factor Socioeconómico. . .	94
5-49.Prueba ajuste modelo Inglés Factor Socioeconómico.	94
5-50.Ajuste de la distribución modelo Inglés Factor Socioeconómico.	95
5-51.Modelo óptimo Lectura Crítica Factor Académico.	97
5-52.Modelo óptimo Lectura Crítica Factor Académico.	98
5-53.Diagnóstico de la distribución modelo Lectura Crítica Factor Académico.	98
5-54.Prueba ajuste modelo Lectura Crítica Factor Académico.	98
5-55.Ajuste de la distribución modelo Lectura Crítica Factor Académico. . .	99
5-56.Modelo óptimo Lectura Crítica Factor Demográfico.	99
5-57.Modelo óptimo Lectura Crítica Factor Demográfico.	100
5-58.Diagnóstico de la distribución modelo Lectura Crítica Factor Demográfico.	101
5-59.Prueba ajuste modelo Lectura Crítica- Factor Demográfico.	101
5-60.Ajuste de la distribución modelo Lectura Crítica Factor Demográfico. .	102
5-61.Valores Atípicos Número de personas a cargo Lectura Crítica.	102
5-62.valores atípicos Edad Lectura Crítica.	102
5-63.valores atípicos Número personas hogar Lectura Crítica.	103
5-64.Prueba de valores atípicos número de personas a cargo Lectura Crítica.	103
5-65.Prueba de valores Atípicos Edad Lectura Crítica.	104
5-66.Prueba de valores atípicos Número de personas en el hogar Lectura Crítica.	104
5-67.Modelo óptimo Lectura Crítica Factor Socioeconómico.	105
5-68.Modelo óptimo Lectura Crítica Factor Socioeconómico.	106
5-69.Diagnóstico de la distribución modelo Lectura Crítica Factor Socioeconómi- co.	107
5-70.Prueba ajuste modelo Lectura Crítica Factor Socioeconómico.	107
5-71.Ajuste de la distribución modelo Lectura Crítica Factor Socioeconómico.	108
5-72.Modelo óptimo Comunicación Escrita Factor Académico.	109
5-73.Diagnóstico de la distribución modelo Comunicación Escrita Factor Académi- co.	110
5-74.Prueba ajuste modelo Comunicación Escrita Factor Académico.	110
5-75.Ajuste de la distribución modelo Comunicación Escrita Factor Académico.	111
5-76.Modelo óptimo Comunicación Escrita Factor Demográfico.	112
5-77.Diagnóstico de la distribución modelo Comunicación Escrita Factor De- mográfico.	113
5-78.Prueba ajuste modelo Comunicación Escrita Factor Demográfico. . . .	113

5-79. Ajuste de la distribución modelo Comunicación Escrita Factor Demográfico.	114
5-80. Valores atípicos Número de personas a cargo Comunicación Escrita. . .	114
5-81. Valores atípicos Edad Comunicación Escrita.	114
5-82. Valores atípicos Número personas hogar Comunicación Escrita.	115
5-83. Prueba de valores atípicos número de personas a cargo Comunicación Escrita.	115
5-84. Prueba de valores Atípicos Edad Comunicación Escrita.	116
5-85. Prueba de valores atípicos Número de personas en el hogar Comunicación Escrita.	116
5-86. Modelo óptimo Comunicación Escrita Factor Socioeconómico.	118
5-87. Modelo óptimo Comunicación Escrita Factor Socioeconómico.	119
5-88. Diagnóstico de la distribución modelo Comunicación Escrita Factor Socioeconómico.	120
5-89. Prueba ajuste modelo Comunicación Escrita Factor Socioeconómico. . .	120
5-90. Ajuste de la distribución modelo Comunicación Escrita Factor Socioeconómico.	121

1 PRESENTACIÓN

El Ministerio de Educación Nacional (MEN) en busca del mejoramiento de la calidad ha implementado en la evaluación de la pruebas Saber Pro un componente genérico que evalúa las competencias de los futuros profesionales de cada centro de educación superior del país, con el fin de establecer un referente de medición externa de la calidad de la educación. Pretendiendo comprobar el grado de desarrollo de las competencias de los estudiantes, programas e instituciones, es por ello que se hace necesario analizar los factores Académicos, Demográficos y Socioeconómicos incidentes en los puntajes de los estudiantes de Licenciatura en Matemáticas y Estadística de la Universidad Pedagógica y Tecnológica de Colombia, facultad seccional Duitama.

En esta investigación se buscó analizar los resultados de las pruebas Saber Pro, presentadas por los estudiantes del programa de Licenciatura en Matemáticas y Estadística, con el fin de identificar los factores que influyen en los resultados del componente genérico, teniendo como fuente las bases de datos suministradas por el ICFES, de tal forma que se establecieron algunos aspectos a mantener o mejorar, los cuales brinden soporte para potenciar las competencias genéricas en los estudiantes del Programa.

La investigación se describe en cinco capítulos, en el primero se presenta el contexto que orientó la investigación, en el segundo se encuentra el marco de referencia sobre pruebas Saber Pro y modelamiento estadístico, en el tercero se presenta la metodología de trabajo, el cuarto muestra el análisis de la información basados en los modelos de regresión logística ordinal (RLO) y modelos aditivos generalizados para localización, escala y forma (GAMLSS) y un quinto capítulo que contiene la presentación de los resultados, finalizando con las conclusiones y discusión.

La prueba Saber Pro, se divide en dos bloques: el primero es el módulo de competencias genéricas y el segundo corresponde a competencias específicas comunes. Estas pruebas constituyen un instrumento estandarizado para la evaluación externa de la calidad de la educación superior, y tienen como principales estándares caracterizar el grado de calidad de la educación y la evaluación (ICFES, 2009), además de medir el desarrollo de las competencias que tienen los futuros profesionales, y que son necesarias para afrontar los desafíos en el mercado laboral, haciendo obligatorio que todos los estudiantes que están próximos a graduarse de los diferentes programas de educación superior presenten la prueba. Para la Organización para la Cooperación y el Desarrollo Económico (OCDE) incluir competencias genéricas de acuerdo con el proyecto, Evaluación de los Resultados del Aprendizaje en Educación Superior (AHELLO), mide competencias en pensamiento crítico, razonamiento analítico, resolución de problemas, facilidad de comunicación escrita, capacidad de liderazgo y capacidad de trabajo en equipo. A esta lista se puede añadir la competencia en una segunda lengua. Asimismo la OCDE y el Banco Mundial, en la revisión realizada en 2012 sobre la educación superior, encontraron que los estudiantes Colombianos están menos preparados en dichas pruebas en comparación con otros países, debido a que nuestro país son más jóvenes y sus estándares educativos son bajos, con excepción de estudiantes de altos ingresos económicos.

La prueba Saber Pro para la educación superior tiene un alto impacto en las universidades, generando cambios en los currículos, formas de evaluación y en la forma como la universidad prepara a los estudiantes para presentar la prueba. Además constituye un referente de medición para la universidad en un ámbito local y nacional, posicionándola a través de dicha prueba, de modo que conocer aquellos factores que inciden en los resultados pueden contribuir a mejorar la calidad académica y genera procesos de auto-evaluación a fin de mejorar en competencias como Razonamiento Cuantitativo, Lectura Crítica, Inglés y Comunicación escrita como lo afirma (MESA, 2009).

A nivel nacional, Castellanos *et al* (2014), desarrollaron un proyecto donde se analizaron los resultados de la pruebas 2010 de los estudiantes de últimos semestres de Ingeniería Civil, con el fin de determinar los factores que influyen en los puntajes obtenidos en la prueba Saber Pro. Se estudió la asociación de variables, cualitativas como cuantitativas, con el puntaje obtenido, mediante tablas de contingencia y el análisis de correspondencias simple y múltiple. Se encontró, al realizar un análisis de dependencia entre las variables del estudio, que no fue relevante la situación laboral, ni el título de bachiller en el puntaje de la prueba. Además los puntajes altos se asocian a los estu-

diantes de género masculino, mientras los bajos al femenino, en donde estos puntajes superiores a la media fueron obtenidos por estudiantes de estratos sociales medio y alto pertenecientes a las instituciones de educación pública.

Un estudio realizado por Linero (2009), enfocado a medir la incidencia en el resultado de las Prueba ECAES de factores académicos tales como los resultados obtenidos de las materias cursadas durante la carrera, los tópicos preparatorios ofrecidos por la Facultad de Economía y la participación de los estudiantes en éstos. En el área de Econometría evaluada en el año 2005 se encontró que los estudiantes que obtuvieron una buena nota en la materia cursada durante la carrera lograron sacar un buen resultado en el componente de econometría y estadística de las pruebas ECAES. Si los estudiantes participaran en los cursos preparatorios ofrecidos por la facultad de Economía en las siguientes materias Microeconomía, Macroeconomía, Historia del pensamiento económico y Econometría, van a obtener unos mejores resultados que los conseguidos en las pruebas ECAES presentadas en los dos años objeto del estudio.

Manrique, J. y Carreño, C. (2014), dan a conocer un artículo con perspectiva econométrica, donde buscan establecer las relaciones, entre el desempeño de los estudiantes en las pruebas Saber Pro 11, con el nivel educativo de los padres de familia, el hecho de tener internet y el tipo de género del estudiante. Se apoyó en una regresión econométrica, permitiendo así conformar dos modelos de regresión logit para establecer el efecto de los factores endógenos y exógenos, en los puntajes obtenidos por los estudiantes en las áreas de matemáticas y lenguaje. Mediante el empleo de dicha técnica estadística, se encontró que el internet es una variable que afecta positivamente el rendimiento académico, dado que al poseer este servicio aumenta la probabilidad de obtener un puntaje superior a 70 en la prueba de lenguaje.

El factor género no es significativo para la prueba de lenguaje, por lo tanto, el ser hombre o mujer no afecta en las probabilidades de sacar más de 70 en esta prueba. Por otro lado, en la prueba de matemáticas sí es significativa la diferencia entre ser hombre y mujer. Respecto al tipo de educación del padre (primaria y secundaria) esta variable no es significativa en ninguna regresión. Una de las causas aparentes de este fenómeno, es la baja influencia o ausencia de los padres dentro de la educación de sus hijos.

Al iniciar el proceso de análisis de los datos suministrados por el ICFES, y utilizar un modelo lineal generalizado (GLM) que consiste en aplicar una serie de procedimientos estadísticos con el fin de conseguir una predicción apropiada al comportamiento de los puntajes, a partir de una función ponderada de una o más variables explicativas (Modelo) como lo dice (Cayuela, 2009), se obtuvo como resultado que los puntajes no pertenecían a la familia exponencial de densidades. Para dar solución a esta situación se emplearon los GAMLSS, los cuales pueden trabajar con numerosas distribuciones que se salen de la familia exponencial de densidades. En cuanto a los resultados en Inglés y Comunicación Escrita medidos en desempeños se aplicó la RLO, determinando la incidencia de las variables independientes en los desempeños por medio del estadístico de Wald, los Odds y los intervalos de confianza.

En la población objeto de análisis se encontró que tienen una edad promedio de 27 años, en su mayoría son hombres, cuyos padres tienen mayoritariamente grado escolar de primaria completa, con una ocupación no remunerada en el caso de las madres y los padres con ocupación de administrador. Predomina el estado civil soltero, con un pago de matrícula menor a 500,000 pesos, pertenecientes a un tipo de bachillerato normalista, con situación del hogar permanente o habitual, donde el promedio de personas en el hogar es de cuatro, caracterizados por ser cabeza de familia, con una persona en promedio a cargo, ubicados en estrato dos en el recibo de energía y sisben, los cuales cuentan con celular, telefono fijo y lavadora, de manera opuesta no cuentan con internet ni televisión. El ingreso familiar esta entre dos y tres salarios mínimos legales vigentes, tienen en promedio tres dormitorios en el hogar, se desempeñan laboralmente como ayudantes sin remuneración trabajando en un promedio de 16 horas por semana, caracterizados por repasar por cuenta propia para la prueba Saber Pro.

Se encontró que la educación de los padres (primaria y secundaria) influye de manera negativa en el puntaje de cada componente, mientras que en los desempeños en Comunicación Escrita e Inglés no se presentó incidencia por estas variables.

Se observó que pertenecer al sexo masculino, ser cabeza de familia, y tener: un número mayor de años, una situación del hogar temporal, un número mayor de personas a cargo y en el hogar, aumenta el puntaje de Razonamiento Cuantitativo.

Para el puntaje de Inglés, se observa que ser casado, tener mayor edad y una situación del hogar temporal, aumenta el puntaje, mientras que ser hombre y tener un número mayor de personas en el hogar lo disminuye. El desempeño en inglés decrece en todos los niveles al tener mayor edad.

Respecto a Lectura Crítica, se encontró que ser cabeza de familia y tener una mayor edad disminuye el puntaje, en cambio tener un número mayor de personas a cargo lo aumenta, como también tener televisión y un salario mensual familiar entre 4 y 5 S.M.V, aumenta el puntaje.

El ser cabeza de familia casado o soltero con mayor edad, y tener un número mayor de personas en el hogar, disminuye el puntaje de Comunicación Escrita, mientras que tener un número mayor de personas a cargo lo aumenta. Contar con servicio de internet aumenta el Desempeño en Comunicación Escrita, mientras que tener teléfono fijo lo disminuye.

Se observó que las variables del Factor Socioeconómico no influyen en los puntajes de Razonamiento Cuantitativo, Comunicación Escrita y el Desempeño en Inglés.

Respecto al puntaje de Inglés, se encontró que pertenecer al estrato dos, donde la ocupación de la madre es obrero, y tener: Celular, un mayor número de dormitorios y horas de trabajo, lavadora, trabajo, un valor de matrícula menor a 500,000 mil y otro de tres a cinco millones de pesos, disminuye el puntaje.

De acuerdo con Castellanos *et al* (2014), existe una diferencia entre los puntajes obtenidos en Inglés por los estudiantes de diferentes estratos, como también una asociación con el desempeño en Comunicación Escrita. Además se encontró una diferencia en los promedios respecto al sexo en Razonamiento Cuantitativo, obteniendo puntajes más altos las mujeres en este componente, de manera opuesta como lo menciona Castellanos *et al* (2014).

Al comparar resultados con Manrique, J. y Carreño, C. (2014), el sexo no afecta el desempeño en Comunicación Escrita, es decir no importa ser hombre o mujer para obtener niveles alto o bajos, ya que no es una variable influyente. Para el caso de internet,

resulta tener una influencia positiva en el desempeño de Comunicación Escrita como concluyeron Manrique, J. y Carreño, C. (2014). De manera distinta con Manrique, J. y Carreño, C. (2014), la educación del padre (primaria y secundaria), si presento incidencia en el promedio del desempeño en Comunicación escrita y en el puntaje de Inglés.

Se recomienda una incorporación de Inglés en el plan de estudios de tal manera que influya en el promedio acumulado, esto sería lo adecuado para fortalecer el desempeño en esta área y así lograr mejores resultados en este componente de la pruebas Saber Pro. Por otro lado, el internet es una herramienta relevante para el desarrollo general de las temáticas, contar con un apto funcionamiento de este en la Universidad, sería de gran ayuda para el desempeño de los estudiantes de Licenciatura en Matemáticas y Estadística.

2 REFERENTES TEÓRICOS

En los siguientes párrafos se abordan las teorías que dan soporte a esta investigación en cuanto a temas de: evaluación en educación, pruebas genéricas, pruebas Saber Pro, factores Académicos, Demográficos, Socioeconómicos, y algunos aspectos del modelamiento estadístico.

2.1 ANTECEDENTES

A nivel nacional, Castellanos *et al* (2014), desarrollaron un proyecto donde se analizaron los resultados de la pruebas 2010 de los estudiantes de últimos semestres de Ingeniería Civil, con el fin de determinar los factores que influyen en los puntajes obtenidos en la prueba Saber Pro. Se estudió la asociación de variables, cualitativas como cuantitativas, con el puntaje obtenido, mediante tablas de contingencia y el análisis de correspondencias simple y múltiple. Se encontró, al realizar un análisis de dependencia entre las variables del estudio, que no fue relevante la situación laboral, ni el título de bachiller en el puntaje de la prueba. Además los puntajes altos se asocian a los estudiantes de género masculino, mientras los bajos al femenino, en donde estos puntajes superiores a la media fueron obtenidos por estudiantes de estratos sociales medio y alto pertenecientes a las instituciones de educación pública.

Un estudio realizado por Linero (2009), enfocado a medir la incidencia en el resultado de las Prueba ECAES de factores académicos tales como los resultados obtenidos de las materias cursadas durante la carrera, los tópicos preparatorios ofrecidos por la Facultad de Economía y la participación de los estudiantes en éstos. En el área de Econometría evaluada en el año 2005 se encontró que los estudiantes que obtuvieron una buena nota en la materia cursada durante la carrera lograron sacar un buen resultado en el componente de Econometría y Estadística de las pruebas ECAES. Si los estu-

diantes participaran en los cursos preparatorios ofrecidos por la facultad de Economía en las siguientes materias Microeconomía, Macroeconomía, Historia del pensamiento económico y Econometría, van a obtener unos mejores resultados que los conseguidos en las pruebas ECAES presentadas en los dos años objeto del estudio.

Manrique, J. y Carreño, C. (2014), dan a conocer un artículo con perspectiva econométrica (modelos matemáticos y estadísticos), donde buscan establecer las relaciones, entre el desempeño de los estudiantes en las pruebas Saber Pro 11, con el nivel educativo de los padres de familia, el hecho de tener internet y el tipo de género del estudiante. Se apoyó en una regresión econométrica, permitiendo así conformar dos modelos de regresión logit para establecer el efecto de los factores endógenos y exógenos, en los puntajes obtenidos por los estudiantes en las áreas de matemáticas y lenguaje. Mediante el empleo de dicha técnica estadística, se encontró que el internet es una variable que afecta positivamente el rendimiento académico, dado que al poseer este servicio aumenta la probabilidad de obtener un puntaje superior a 70 en la prueba de lenguaje. El factor género no es significativo para la prueba de lenguaje, por lo tanto, el ser hombre o mujer no afecta en las probabilidades de sacar más de 70 en esta prueba. Por otro lado, en la prueba de matemáticas sí es significativa la diferencia entre ser hombre y mujer. Respecto al tipo de educación del padre (primaria y secundaria) esta variable no es significativa en ningún puntaje. Una de las causas aparentes de este fenómeno, es la baja influencia o ausencia de los padres dentro de la educación de sus hijos.

2.2 EL CONTEXTO DE LA EVALUACIÓN

Según Tolosa *et al* (2014), los organismos a nivel internacional han tenido cierta influencia importante en temas políticos internos trascendentales en la educación superior, como lo son el Banco Mundial, el Banco Interamericano de Desarrollo (BID), la Organización para la Cooperación y el Desarrollo Económico (OCDE) y el Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL). En 1944 después de la segunda guerra mundial surge el Banco Mundial el cual es un organismo prestamista, perteneciente a Naciones Unidas teniendo como función realizar estos préstamos del interés moderado apoyando en procesos financieros. Los estamentos del Banco Mundial son: el BIRF (Banco Internacional de reconstrucción y fomento), la AIF (Asociación Interna de Fomento), la CFI (Corporación Financiera Internacional), el OMGI (Organismo Multilateral de Garantía de Inversiones) y el CIADI (Centro In-

ternacional de Arreglo de Diferencias Relativas a Inversiones), donde el BIRF y la AIF son los que tienen alguna relación de influencia directa con Colombia. De acuerdo a la OCDE¹:

1. Colombia es el quinto país más grande en Latinoamérica con 47.6 millones de habitantes ocupa el tercer puesto en relación a los habitantes, Brasil y México son los líderes poblacionales. Las raíces colombianas son una mezcla de personas provenientes de África, Europa y Oriente Medio. Donde el 76 % de los colombianos reside en zonas urbanas, aunque estos indicadores varían significativamente en relación al país. Los niveles de desarrollo se centran en Bogotá, Medellín, Cali y Barranquilla, estas ciudades con poblaciones superiores a 1 millón de habitantes. Las zonas rurales son las principales donde la infraestructura, seguridad, pobreza y desarrollo son problemas frecuentes.

2. La corrupción sigue como un problema que a medida del tiempo aumenta por medio de una burocracia ineficiente a las necesidades de un país, es un factor incidente en todos los aspectos la educación. Se han realizado procesos permitiendo fortalecer y promover una mayor gestión, pero todavía hay mucho campo para abarcar. Diversos países latinoamericanos presentan la misma situación de Colombia, la cual es tener un sector público pequeño, siendo un aspecto directamente influyente para su labor como tal. Se han determinado instituciones de auditoria con el fin de erradicar la corrupción de Colombia y así mejorar en el aspecto democrático; pero por causa de un sector público pequeño se generan casos de deficiencia en los procesos, generando así la situación de desigualdad regional y territorial por falta de transparencia.

3. El conflicto colombiano existe hace más de cinco décadas, una huella imborrable para la educación; no todos los departamentos de Colombia son atacados directamente por esta problemática; los más afectados son: Arauca, Choco, Nariño, Norte de Santander y Putumayo. Las cifras de homicidio se han disminuido significativamente, pero en comparación con los países que constituyen la OCDE sobresale por una diferencia relevante. Un 25 % de los municipios tuvo índices altos de violencia en el año 2013, desatando la falta de bienestar de las comunidades e instituciones. Las poblaciones que viven internamente en el conflicto están propensas a muerte, reclutamiento forzado, secuestro y exclusión de la educación. La educación se ve limitada para aquellas personas que son directamente atacadas por el conflicto colombiano, personas que en su mayoría

¹OCDE. (2016), p 20,22,23,24. Revisión de políticas nacionales de educación: Educación en colombia

son indígenas o afrocolombianos presentan dicho suceso, se calcula que 5,7 millones son los afectados por el desplazamiento forzado. 110.000 estudiantes en el 2013 se registraron como víctimas del desplazamiento, siendo esta una cifra mayor a la del 2005 con un aumento mayor a cuatro veces.

Mediante el contexto hablado anteriormente se dio a conocer aspectos directamente influyentes en el ambiente educativo y económico de los colombianos, teniendo como implicación diferentes problemáticas hacia los estudiantes que presentan las pruebas Saber Pro, es decir los resultados no solo vienen ligados a los conocimientos del estudiante sino también a su entorno. Es precisamente este aspecto el que se abordará en esta investigación, entorno que será medido parcialmente a través de aspectos académicos, demográficos y socioeconómicos.

Para Perlaza *et al* (2009), desde los años 60 se tiene la idea de velar cada día por la calidad de los egresados de los programas de educación superior. En 1966 se recomendó al Plan Nacional para la Educación Superior en Colombia realizar pruebas a los profesionales graduados con la intención de generar acreditación en las facultades, transferencia de estudiantes y selección de candidatos para posgrados.

Según Rodríguez *et al* (2005), en la década de los 80s se trabajó en proyectos en pro de la evaluación de la calidad de aprendizaje de los egresados; en el periodo de César Gaviria (1990-1994) se ejecutó la idea de llevar a cabo los exámenes de educación superior para carreras de Medicina, Derecho y Contaduría. En el año de 1998 se implementaron los exámenes de educación superior para las ingenierías después de un proceso con diversas organizaciones de profesionales las primeras pruebas fueron para Ingeniería Mecánica fueron ajustadas de acuerdo a un sondeo experimental con egresados y alumnos del último año en el 2000.

Según Rodríguez *et al* (2005), tiempo después se crearon los decretos 1716 y 2233 de 2001 y 1373 de 2002 donde se hacía pertinente los exámenes de educación superior para carreras de Medicina, Ingeniería Mecánica y Derecho. Éste sería el comienzo del proceso de reglamentación para los exámenes de evaluación de calidad en la educación superior en Colombia. considerados en el 2000-2002 en el plan estratégico de educación. De acuerdo a la reglamentación del Gobierno Nacional programas académicos de Ingenierías, Ciencias de la Salud, Derecho, Arquitectura, Comunicación e Información,

Administración, Contaduría Pública, Economía, Psicología y Ciencias Exactas y Naturales serían partícipes de la reglamentación como ya lo venían haciendo las carreras anteriormente nombradas. El ICFES realizó un llamado a las diferentes universidades, asociaciones de profesionales, facultades y demás organizaciones académicas, para que fueran partícipes con sus propuestas de construcción y diseño de los nuevos ECAES; como resultado, se dio una amplia participación de las organizaciones invitadas y se elaboraron 27 pruebas que fueron implementadas el 1 de noviembre de 2003 en 41 ciudades del país y 12 extranjeras, a un total de 58.901 personas, entre estudiantes de último año y egresados de 27 programas de pregrado.

De acuerdo a Tovar (2013), las Instituciones de Educación Superior se encuentran manejadas por el marco de la ley 30 de 1992, definiendo la educación superior como un servicio público, inherente a las metas del estado, no se ve como un bien público a la comunidad sino como un servicio accesible para quienes demuestren poseer las capacidades requeridas y cumplan las condiciones académicas exigidas en cada caso.

De acuerdo al ICFES a continuación se muestra la evolución del examen Saber Pro a través de los años ² :

1. Con la ejecución del decreto 3963 de octubre de 2009, el ICFES vive una constante transformación de los exámenes de estado de calidad de la educación superior, ya que se aplicará en un periodo transitorio. Mientras se genera una nueva perspectiva de las pruebas, esto tiene un tiempo de vigencia mayor igual a 12 años.

2. Por medio del requisito como tal de presentar las pruebas de educación superior por los programas existentes, el ICFES puso en desarrollo la implementación de las pruebas genéricas requeridas por el Ministerio de Educación Nacional (MEN), sin importar el área específica de formación. En 2008 se dieron ejecutadas las exigencias del MEN donde se implementó una prueba piloto en los años 2008-2009 a 1.084 programas cuyos estudiantes fueron evaluados con estas pruebas. Este examen conformado por las Competencias Genéricas, Inglés y Comprensión Lectora, teniendo como un modelo de evaluación para los profesionales futuros en todas las áreas del conocimiento, también el decreto nos interpone la implementación de exámenes específicos donde se evalúen

²Guías, ICFES (2011), p 10-11. Orientaciones para el examen de Estado de calidad de la educación superior SABER PRO (ECAES). Bogotá, DC, septiembre.

los núcleos relevantes de formación.

3. Luego de una aplicación de las pruebas por ocho años para los diversos programas, en el 2010 implementó una reforma a las pruebas de educación superior por medio del acompañamiento de comunidades académicas de áreas como Ingeniería, Salud, Educación, Ciencias Naturales y exactas, Ciencias Agropecuarias y Ciencias Sociales y Humanas, para verificar los marcos teóricos de referencia de las pruebas y así generar nuevas especificaciones de pruebas de competencias comunes de área o programas. Entorno a estas ideas el ICFES le solicitó a ASCOFADE (Asociación Colombiana de Facultades de Educación) la realización de pruebas para el área de educación se dio por medio de todas las facultades de educación de todo el país. Las especificaciones fueron dadas a conocer y validadas con el apoyo de la academia nacional por medio de talleres regionales efectuados en los primeros meses del año 2011. De donde el ICFES ofreció un examen específico para las distintas licenciaturas aplicándose en el 2011. Es de advertir que los resultados de este nuevo examen no podrán compararse con los resultados de años anteriores, debido a que se ha configurado una nueva estructura de prueba.

4. La Ley 1324 de 2009 y el Decreto 3963 del mismo año precisa que el Gobierno nacional reglamentará la implementación gradual de los Saber Pro en los términos de la presente ley.

De esta manera se implementaron las pruebas Saber Pro en el País.

2.2.1. Objetivos de los exámenes - Saber Pro

De acuerdo con el Decreto 3963 de octubre de 2009, son objetivos del Saber Pro los siguientes:

- A. Comprobar el grado de desarrollo de las competencias de los estudiantes próximos a culminar los programas académicos de pregrado que ofrecen las instituciones de educación superior.
- B. Producir indicadores de valor agregado de la educación superior en relación con el nivel de competencias de quienes ingresan a este nivel; proporcionar información para la comparación entre programas, instituciones y metodologías, y mostrar su evolución en el tiempo.
- C. Servir de fuente de información para la construcción de indicadores de evaluación de la calidad de los programas e instituciones de educación superior y del servicio público educativo, que fomenten la cualificación de los procesos institucionales y la formulación de políticas,

y soporten el proceso de toma de decisiones en todos los órdenes y componentes del sistema educativo ³.

Las pruebas Saber Pro se deben presentar de manera obligatoria, por todos los estudiantes que hayan aprobado por lo menos el 75 % de los créditos académicos del Programa correspondiente o que tengan previsto graduarse en el año siguiente a la fecha de aplicación del examen. Las instituciones de educación superior tienen la responsabilidad de reportar ante el ICFES a la totalidad de sus estudiantes que deban presentar el Saber Pro. Cada uno de los estudiantes reportados deberá realizar el proceso de inscripción directamente o a través de la respectiva institución educativa y presentarse a la prueba, de acuerdo con los procedimientos que se establezcan.

2.2.2. Competencias genéricas

Las competencias genéricas son las que se consideran como las que toda persona, independiente de su formación profesional, debería desarrollar, inclusive independientemente de su contexto o realidad social, lo cual establece una preparación para la inserción al mundo laboral y para las exigencias actuales del sistema productivo.

De acuerdo con lo expuesto en el año 2009 se realizó desde el Ministerio de Educación la propuesta para reformular las competencias genéricas para ser evaluadas en la educación superior, la cual concluyó que las nuevas competencias serían:

1. Comunicación en Lengua materna y en otra lengua internacional
2. Pensamiento matemático
3. Cultura científica, tecnológica y gestión de la información
4. Ciudadanía

Con el fin de dar a conocer lo que el ICFES actualmente está evaluando y cómo lo está haciendo, se explica en qué consisten las competencias genéricas y específicas para los futuros licenciados.

³Guías, ICFES (2011), p 10-11. Orientaciones para el examen de Estado de calidad de la educación superior SABER PRO (ECAES). Bogotá, DC, septiembre.

RAZONAMIENTO CUANTITATIVO

Este módulo evalúa la comprensión de conceptos básicos de las matemáticas para analizar, modelar y resolver problemas, aplicando métodos, procedimientos cuantitativos y esquemáticos que de acuerdo con el ICFES los encargados de hacer la elaboración de preguntas son profesionales formados en el área de Ingenierías o carreras afines. Las competencias a tener en cuenta en este módulo son ⁴:

Interpretar datos

Consiste en la comprensión e interpretación de datos presentados de diferentes formas (tablas, gráficas, esquemas, símbolos, expresiones verbales), así como la generación de diversas representaciones a partir de datos dados.

Formular y ejecutar

Involucra procesos relacionados con la identificación del problema, la construcción y proposición de estrategias adecuadas para su solución además del tratamiento de datos, la modelación y el uso de herramientas cuantitativas (aritméticas, métricas, geométricas, algebraicas elementales y de probabilidad y estadística).

Evaluar y Validar

Incluye procesos relacionados con la verificación de resultados, hipótesis o conclusiones que se derivan de la interpretación y de la modelación de situaciones.

LECTURA CRÍTICA

Evalúa competencias relacionadas con la capacidad para leer de manera analítica y reflexiva, lo que exige al lector identificar la información presente en uno o varios textos, construyendo así su sentido global, estableciendo relaciones entre enunciados y evaluar su intencionalidad. En el módulo se abordan procesos relacionados con ⁵ :

⁴Tolosa, F., Fernanda, L., Rubio Mendoza, H. A., & Rojas Quevedo, D. G. (2014) , p 46. El papel de la evaluación de competencias en los procesos de formación docente: análisis de las pruebas Saber Pro para Licenciados de Psicología y Pedagogía, Licenciados en Español y Lenguas Extranjeras en la Universidad Pedagógica Nacional.

⁵Ibid.,p.47

Evalúa competencias relacionadas con la capacidad para leer de manera analítica y reflexiva, lo que exige al lector identificar la información presente en uno o varios textos, construyendo así su sentido global, estableciendo relaciones entre enunciados y evaluar su intencionalidad. En el módulo se abordan procesos relacionados con:

Ubicar información

Es la capacidad de encontrar información específica en los textos según un criterio de búsqueda dado. La información que se pide hallar puede ser más o menos evidente y solicitarse empleando las mismas expresiones que aparecen en el texto o utilizando otras equivalentes.

Relacionar

Es la identificación de las relaciones entre distintas partes de los textos. Las relaciones pueden ser de implicación, inclusión, pertenencia, causalidad, orden, ejemplificación, categorización, equivalencia, complementariedad, oposición, contradicción y/o contraste, analogía o contra argumentación.

Construir la representación global del texto

Se refiere a la comprensión del texto como un todo y la construcción del sentido global a partir de la interpretación de sus componentes implícitos y explícitos.

Analizar y evaluar la relación entre procesos discursivos y contexto socio-cultural

Es la capacidad para tomar distancia del texto y rastrear las concepciones de mundo subyacentes, mediante la identificación de las estrategias discursivas utilizadas y el conocimiento del rol de quienes participan en la materialización de los discursos.

INGLÉS

Evalúa la competencia para comunicarse efectivamente en inglés, dichas competencias, alineadas con el Marco Común Europeo, permiten clasificar a los examinados en cuatro niveles de desempeño A1, A2, B1, B2 ⁶ .

⁶Ibid.,p.47

A1

Comprender y utilizar expresiones cotidianas de uso muy frecuente, así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato. Presentarse a sí mismo y a otros, pedir y dar información básica sobre domicilio, pertenencias y personas conocidas. Relacionarse de forma elemental siempre que su interlocutor hable despacio y con claridad.

A2

Comunicarse cuando se llevan a cabo tareas simples y cotidianas que no requieran más que intercambios sencillos y directos de información sobre cuestiones conocidas o habituales. Describir en términos sencillos aspectos del pasado y del entorno, así como cuestiones relacionadas con las necesidades inmediatas del estudiante.

B1

Comprender los contenidos principales de textos claros y en lengua estándar, si tratan cuestiones conocidas. Producir textos sencillos y coherentes sobre temas que son familiares al estudiante o en los que tiene un interés personal. Describir experiencias, acontecimientos, deseos y aspiraciones, así como justificar brevemente sus opiniones o explicar sus planes.

B2

Entender las ideas principales de textos complejos que traten de temas concretos y abstractos, incluso son de carácter técnico, siempre que estén dentro de su especialización. Relacionarse con hablantes nativos con un grado suficiente de fluidez y naturalidad, de modo que la comunicación se realice sin esfuerzo por ninguno de los interlocutores. Producir textos claros y detallados en torno a temas diversos.

ESCRITURA

Este módulo de acuerdo a Tolosa *et al* (2014), evalúa la competencia para comunicar ideas por escrito referidas a un tema dado, que según la entrevista aplicada al ICFES (Tolosa) , son temas de actualidad, que conmuevan la conciencia e inciten un ejercicio escritural, por lo tanto lo que se plantea es que cada estudiante debe elaborar un escrito argumentativo. El ICFES al evaluar este módulo, reconoce a las siguientes acciones

como fundamentales en un buen ejercicio de escritura ⁷:

Expresar adecuadamente la intención comunicativa.

Dar coherencia y cohesión al texto.

Expresarse mediante un lenguaje apropiado, aplicando las reglas que rigen el lenguaje escrito.

Para medir el nivel de desempeño, se estableció desde el ICFES una escala, en la que se encuentran ocho niveles, el cuál de acuerdo con el desempeño mostrado se tipifica de la siguiente manera :

Nivel 8

El texto trasciende el estímulo dado, complejizando los planteamientos de manera efectiva tanto a nivel de pensamiento como de recursos lingísticos. Incluye el problema planteado en un diálogo de ideas y posiciones, en una perspectiva más amplia e intertextual que construye y precisa su sentido. El texto de este nivel es de alto interés y motivación para el lector, debido a la eficacia del escrito.

Nivel 7

En el texto se evidencia una planeación en la escritura en dos aspectos fundamentales. En el primero, el autor piensa en cómo expresar sus ideas de manera efectiva, aplicando diversos recursos textuales para evidenciar sus planteamientos. En el segundo, el autor adecua su discurso para un público determinado, prevé los conocimientos previos de su lector y busca el lenguaje y los conceptos apropiados.

Nivel 6

En el texto se identifica la posición de quien escribe, se expresan con claridad las ideas y hay un uso correcto de las expresiones que permiten conectarlas. Hay un uso adecuado de distintos mecanismos que le dan coherencia y cohesión al texto (signos de puntuación, conectores, etc.). Lo anterior permite que el escrito sea fluido.

⁷Ibid.,p.49

Nivel 5

El texto alcanza unidad, por medio de la progresión temática, es decir, logra encadenar o relacionar efectivamente las ideas, dándole continuidad al escrito: incorpora información nueva vinculándola con la anterior, presentando la información en un orden cronológico, partiendo de temas generales para desglosar temas específicos, entre otras maneras de interrelacionar contenidos. Todo el texto se desarrolla en un mismo eje temático, aunque pueden presentarse fallas en el uso de conectores. Hay un buen uso del lenguaje, aunque pueden encontrarse errores en la aplicación de algunas reglas de ortografía y puntuación.

Nivel 4

En el texto se encuentra una idea central que se desarrolla de acuerdo con una intención comunicativa. El texto también posee una estructura básica, en otras palabras, es posible identificar una introducción al tema que se abordará, un desarrollo y una conclusión. Sin embargo, el texto no incluye toda la información necesaria (progresión temática), su organización no es completamente efectiva, o rompe la unidad al incluir temas que no se relacionan con el marco semántico que desarrolla. Se aprecia un uso aceptable del lenguaje (se aplican las reglas gramaticales más importantes).

Nivel 3

En el escrito se aprecia el esbozo de una intención comunicativa, es decir, se ve que quien escribe pretende alcanzar un fin, por medio del escrito busca una reacción específica en el lector. Se encuentran problemas de manejo del lenguaje, pero estos no impiden la comprensión de los enunciados.

Nivel 2

En el escrito se aprecian ideas, pero estas pueden ser incoherentes, o se presentan desarticuladamente. No hubo desarrollo organizado del tema o el escrito pudo ser innecesariamente largo o repetitivo.

Nivel 1

En el escrito se aborda la tarea propuesta, pero hay problemas en el manejo de las convenciones (serios errores de sintaxis, puntuación o escritura de las palabras) o un desarrollo insuficiente del tema (es tan breve que no se pueden aplicar los criterios de análisis).

COMPETENCIAS CIUDADANAS

Este módulo evalúa las competencias de los estudiantes para analizar, comprender su entorno, en el marco ético que inspira la Constitución Política de Colombia, se evalúan la comprensión de conceptos básicos de la Constitución y competencias como la valoración de argumentos, la multiperspectividad y pensamiento sistémico. Las dimensiones a tener en cuenta para el proceso evaluativo son ⁸:

Comprender qué es la Constitución Política de Colombia y sus fundamentos

Conocer las características básicas de una Constitución. Reconocer que la Constitución promueve la diversidad étnica y cultural del país, y establece que es deber del Estado su protección. Conocer que Colombia es un estado social de derecho e identifica sus características.

Conocer los derechos y deberes que la Constitución consagra

Conocer los derechos fundamentales de los individuos. Conocer los derechos sociales, económicos y culturales.

Conocer los derechos colectivos y del ambiente.

Conocer los deberes de los ciudadanos que consagra la Constitución.

Conocer la organización del Estado de acuerdo con la Constitución

Conocer las funciones y alcances de las ramas del poder y de los organismos de control. Examinar los argumentos presentados.

Valorar la solidez y pertinencia de la argumentación.

Reconocer, contrastar y valorar distintas perspectivas

Analizar situaciones de conflicto.

Reconocer diferencias de concepciones o ideologías.

Analizar alternativas de solución para conflictos o problemas sociales.

Comprender que los problemas o fenómenos sociales y sus soluciones son multidimensionales y reconocer interacciones entre estas dimensiones

Caracterizar y contrastar las dimensiones que hacen parte de una situación, de un problema o de su solución.

⁸Ibid.,p.52

Reconocer y comprender interacciones entre distintas dimensiones de problemas sociales.

Reconocer y comprender las consecuencias que una propuesta de solución tiene en diferentes dimensiones de un problema social.

Caracterizar y contrastar las dimensiones que hacen parte de una situación, un problema o su solución.

Analizar y evaluar la pertinencia y solidez de enunciados

Examinar los argumentos presentados.

Valorar la solidez y pertinencia de la argumentación.

En el caso del componente de profundización de las pruebas Saber Pro, para el caso específico en educación puede verse en el sitio web www.icfes.gov.co.

Como se observó anteriormente los desempeños presentan una descripción cualitativa, en cuanto a los componentes de Razonamiento Cuantitativo, Lectura Crítica, Comunicación Escrita e Inglés, se muestra la siguiente tabla la cual ilustra la escala de interpretación de resultados de las pruebas Saber Pro de la siguiente forma: ⁹.

PUNTAJE	RESULTADO
Menor a 9,0	Muy bajo
Entre 9.0 y 9.5	Bajo
Entre 9.6 y 10.2	Regular
Entre 10.3 y 10.6	Bueno
Entre 10.7 y 11	Muy bueno
Mayor a 11,0	Excelente

2.3 MODELOS LINEALES GENERALIZADOS (GLM)

Los modelos lineales se basan en los siguientes supuestos: Los errores se distribuyen normalmente, la varianza es constante y la variable respuesta se relaciona linealmente con la(s) variable(s) independiente(s), Cayuela, L. (2009).

⁹Sánchez, Ivan de Jesus. Guía de interpretación de resultados de las pruebas Saber Pro. 2013

Para mejorar la flexibilidad de la variable respuesta del modelo LM (modelo lineal) aparecen los GLM. Las flexibilizaciones que se realizaron puntualmente sobre los GLM no fueron suficientes y dieron lugar a la aparición de los modelos GAM (modelos aditivos generalizados). Estos últimos permiten la utilización de funciones suavizadoras para describir la relación entre la variable respuesta y las variables explicativas de un solo parámetro. No obstante, ya a pesar de las modificaciones sobre los GAM, todos estos modelos quedan obsoletos para aquellas distribuciones que tienen más de un parámetro (que son la mayoría de ellas). Los modelos mencionados ahora son capaces de modelar el parámetro para localización directamente de las covariables y en algunas ocasiones se logra modelar el parámetro de escala pero siempre en relación con el parámetro para localización. Por esta razón aparecen los GAMLSS (Modelos Aditivos Generalizados para localización, Escala y Forma). Los modelos permiten la modelización del parámetro de escala, e incluso de los parámetros de forma (asimetría y curtosis) de la distribución utilizada directamente a partir de las variables explicativas¹⁰.

Los modelos lineales generalizados (GLM de las siglas en inglés de Generalized Linear Models) son una extensión de los modelos lineales que permiten utilizar distribuciones no normales de los errores (binomiales, Poisson, gamma, entre otros) y varianzas no constantes. Los GLM tienen dos propiedades importantes Cayuela, L. (2009), las cuales son:

La estructura de los errores

Los errores pueden tener estructura poisson, son útiles para conteos; tipo binomiales, son de utilidad para proporciones y datos de presencia o ausencia; los tipo gamma, útiles con datos que muestran un coeficiente de variación constante y los tipo exponenciales, útiles para los análisis de supervivencia.

La función enlace

Otra razón por la que un modelo lineal puede no ser adecuado para describir un fenómeno determinado es que la relación entre la variable respuesta y la(s) variable(s) independiente(s) no es siempre lineal. La función de vínculo o enlace, por tanto, se

¹⁰Pena, M. J. P. Modelos Aditivos Generalizados para Localización, Escala e Forma (GAMLSS).

encarga de linealizar la relación entre la variable respuesta y la(s) variable(s) independiente(s) mediante la transformación de la variable respuesta. Cayuela, L. (2009). En la figura 2-1 se presentan las funciones de enlace más utilizadas.

Según Cayuela, L (2009), se denominan funciones de enlace canónicas a las funciones que se aplican por defecto a cada una de las distribuciones de errores. Esto no significa que siempre se deba usar una única función de enlace para una determinada distribución. De hecho, puede ser recomendable comparar diferentes funciones de enlace para un mismo modelo y ver con cuál se obtiene un mejor ajuste del modelo a los datos.

Figura 2-1: Algunas funciones de enlace

Función enlace	Fórmula	Uso
Identidad	μ	Datos continuos con errores normales (regresión y ANOVA)
Logarítmica	$\log(\mu)$	Conteos con errores de tipo poisson
Logit	$\log\left(\frac{\mu}{n-\mu}\right)$	Proporciones (datos entre 0 y 1) con errores binomiales
Recíproca	$\frac{1}{\mu}$	Datos continuos con errores gamma
Raíz cuadrada	$\sqrt{\mu}$	Conteos
Exponencial	μ^n	Funciones de potencia

Fuente: Cayuela, L. (2009), p 68

2.4 MODELOS ADITIVOS GENERALIZADOS PARA LOCALIZACIÓN, ESCALA Y FORMA (GAMLSS)

De acuerdo a Peña (2015-2016), “el modelo lineal generalizado (GLM) y el modelo aditivo generalizado (GAM) ocupan un lugar importante en el campo de las técnicas estadísticas para la regresión invariante, véase en Nelder y Wedderburn (1972) y Hastie y Tibshirani (1990) respectivamente”. Estos modelos consideran trabajar con una variable respuesta donde su distribución pertenece a la familia exponencial, modelando la media a partir de las variables explicativas y su varianza respecto a la media, dada por la siguiente expresión $Var(y) = \phi(u)$, donde ϕ es un parámetro de dispersión constante. Además, si se considera una distribución de la familia exponencial, tanto

la asimetría como la curtosis están en función de la media ϕ del parámetro (las estimaciones de los parámetros dependen totalmente de la estimación de la media con relación a las variables explicativas). Tanto si se trabaja con modelos GLM o GAM ni la varianza, ni la asimetría, ni la curtosis son modeladas explícitamente a partir de las variables explicativas solo a través de la relación de variables independientes con la media.

Los modelos aditivos generalizados para localización, escala y forma (GAMLSS) fueron introducidos por Rigby y Stasinopoulos (2001, 2005) y Akantziliotou et al. (2002) como un punto para solucionar algunas de las limitaciones que presentaban los modelos lineales generalizados (GLM) y los Modelos aditivos generalizados (GAM).

Los modelos GAMLSS son una clase de modelos univariantes, se pueden considerar como modelos de regresión semiparamétricos. Son paramétricos en el sentido como se distribuye la variable respuesta y semi en el sentido en el que el modelado de los parámetros de la distribución de las variables explicativas puede implicar el uso de funciones suavizadas no paramétricas.

En los modelos GAMLSS la variable respuesta y se flexibiliza, permitiendo trabajar con distribuciones con alta asimetría o curtosis. Por lo tanto, la variable respuesta puede ser seleccionada entre una amplia gama de distribuciones tanto continuas como discretas.

Además, la parte sistemática del modelo es amplia permitiendo modelar no solo a la media (parámetro para localización), sino también en el caso de la varianza (parámetro de escala), la asimetría o la curtosis (parámetros de forma). Incluyendo relaciones lineales y no lineales entre los parámetros y las distintas variables explicativas.

El modelo GAMLSS modela no solo la media μ , sino también el parámetro σ de escala (Varianza) y los parámetros de forma ν y τ (asimetría y curtosis), el modelo que resulta es ¹¹:

$$g_1(\mu) = X_1\beta_1 + \sum_{j=1}^{j_1} h_{j_1}(x_{j_1}) \quad (2-1)$$

¹¹Peña, M. J. P. Modelos Aditivos Xeneralizados para Localización, Escala e Forma (GAMLSS).

$$g_2(\sigma) = X_2\beta_2 + \sum_{j=1}^{j_2} h j_2(x_{j2}) \quad (2-2)$$

$$g_3(\nu) = X_3\beta_3 + \sum_{j=1}^{j_3} h j_3(x_{j3}) \quad (2-3)$$

$$g_4(\tau) = X_4\beta_4 + \sum_{j=1}^{j_4} h j_4(x_{j4}) \quad (2-4)$$

De manera generalizada se tiene

$$g_k(\theta_k) = X_k\beta_k + \sum_{j=1}^{j_k} h j_k(x_{jk}) \quad (2-5)$$

Donde θ_k es el vector de parámetros de la distribución.

En GAMLSS la estimación se lleva a cabo maximizando la siguiente función de verosimilitud penalizada (restringida para ciertos valores):

$$l_p = l - \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^{j_k} \lambda_{jk} \Upsilon'_{jk} G_{jk} \Upsilon_{jk} \quad (2-6)$$

Donde $l = \sum_{i=1}^n \log f(y_i|\theta_i)$ es el logaritmo de la función de verosimilitud λ_{jk} son los parámetros de penalización de G_{jk} de una matriz simétrica que depende de los parámetros λ_{jk} .

Para el ajuste de los modelos GAMLSS existen dos algoritmos : RS que se basa en el algoritmo descrito en Rigby y Stasinopoulos (1996a) y CG que se basa en el algoritmo de Cole y Green (1992). Un modelo GAMLSS se fundamenta en 4 componentes fundamentales $M = \{D, G, T, \Lambda\}$, donde

- D especifica la distribución de la variable respuesta.
- G especifica un conjunto de funciones de enlace (g_p, \dots, g_p) para los parámetros $(\theta_1, \dots, \theta_p)$.

- T abarca las variables utilizadas en los predictores correspondientes μ, σ, ν y τ .
- Λ abarca los diversos parámetros de suavizado para las funciones de suavizado.

Para un conjunto de datos, en particular, el proceso de selección es comparar muchos modelos contruidos combinando los diferentes componentes de M . Hablemos ahora de cada uno de estos componentes por separado:

- Componente D :
Una selección de una distribución apropiada esta dada por dos pasos:El primer paso es el aislar diferentes modelos utilizando diferentes distribuciones con la que proporcione un modelo con menor GAIC (Criterio Generalizado de Información de Akaike). El siguiente paso es validar dicha elección a través de un proceso de diagnóstico como puede ser el worm plot.Existen una gran variedad de familias relacionadas con GAMLSS con distribuciones continuas y discretas, como tambien paquetes como `gamlss.tr` y `gamlss.cens` permitiendo construir distribuciones a partir de las existentes.
- Componente G :
La selección de la función de enlace viene determinada generalmente por el rango de valores de la variable explicativa. Una buena elección de función de enlace puede mejorar el ajuste del modelo considerablemente; la elección de esta función será sobre el uso del criterio de la menor deviance (generalización del análisis de varianza utilizado para los GLM).
- Componente T :
La selección de los términos aditivos sobre el modelo se puede hacer es a través del proceso de avance, o por pasos hacia atrás. Además, estos procesos pueden ser aplicadás a cada parámetro por separado o sobre todos los parámetros a la vez.
- Componente Λ :
Para cada terminación suavizada necesitaremos su correspondiente parámetro de suavizado. Este último puede ser previamente fijado o estimado a partir de los datos, es más aconsejable estimar el parámetro de suavizado automáticamente.

2.4.1. Algunas distribuciones asumidas en los GAMLSS

A continuación algunos tipos de distribuciones seleccionadas para el ajuste de los punajes del componente genérico ¹²

Distribución Reverse Gumbel (RG)

La distribución inversa de Gumbel, que también se denomina distribución de valores extremos tipo I, es un caso especial de distribución de valores extremos generalizada [véase Johnson et al. (1995) p75]. La distribución inversa de Gumbel es apropiada para una asimetría moderadamente positiva datos. Esta definida como

$$f_Y(y|\mu, \sigma) = \frac{1}{\sigma} \exp \left\{ - \left(\frac{y - \mu}{\sigma} \right) - \exp \left[\frac{-(y - \mu)}{\sigma} \right] \right\} \quad (2-7)$$

Para $-\infty < y < \infty$, donde $-\infty < \mu < \infty$ y $\sigma < 0$, con $E(Y) = \mu + \gamma\sigma \simeq \mu + 0,57722\sigma$ y $Var(Y) = \frac{\pi^2\sigma^2}{6} \simeq 1,64493\sigma^2$. Nota: si $Y \sim RG(\mu, \sigma)$ y $W = -Y$, entonces $W \sim GU(-\mu, \sigma)$.

Distribución Skew Normal 2 (SN2)

Esta distribución SN2 (μ, σ, ν) se denota

$$f_Y(y|\mu, \sigma, \nu) = \frac{c}{\sigma} \left\{ \exp \left[-\frac{1}{2}|\nu z|^2 \right] I(y < \mu) + \exp \left[-\frac{1}{2}\left|\frac{z}{\nu}\right|^2 \right] I(y \geq \mu) \right\} \quad (2-8)$$

Para $-\infty < y < \infty$, donde $-\infty < \mu < \infty$, $\sigma > 0$ y $\nu > 0$ y donde $z = (y - \mu)/\sigma$ y $c = 2\nu / [(1 + \nu^2)\sqrt{2}\Gamma(\frac{1}{2})]$ Fernandez, Osiewalski and Steel (1995). Nota este $I()$ es un indicador de la función, donde $I(\mu) = 1$ si μ es verdadero y $I(\mu) = 0$ si μ es falso.

Obsérvese que μ es la media de Y . Aquí $E(Y) = \mu + \sigma E(z)$ y $Var(Y) = \sigma^2 V(Z)$ donde $E(Z) = \sqrt{2}\Gamma(\nu - \frac{1}{\nu})/\Gamma(\frac{1}{2})$ y $E(Z^2) = 2\Gamma(\frac{3}{2})(\nu^3 + \frac{1}{\nu^3})/\Gamma(\frac{1}{2})(\nu + \frac{1}{\nu})$, Fernandez, Osiewalski y Steel (1995), p1333, ecuaciones. (12) y (13). La distribución SN2 (μ, σ, ν) es un caso especial de la distribución SEP3 (μ, σ, ν, τ) donde $\tau = 2$, Johnson et al. (1994), p173.

¹²Stasinopoulos, M., Rigby, B., & Akantziliotou, C. (2008). Instructions on how to use the gamlss package in R Second Edition.

Distribución Exponential Gaussian (exGAUS)

ExGAUS (μ, σ) , dada por

$$f_Y(y|\mu, \sigma, \nu) = \frac{1}{\nu} \exp \left[\frac{\mu - y}{\nu} + \frac{\sigma^2}{2\nu^2} \right] \Phi \left(\frac{y - \mu}{\sigma} - \frac{\sigma}{\nu} \right) \quad (2-9)$$

Para $-\infty < y < \infty$, donde $-\infty < \mu < \infty$, $\sigma > 0$ y $\nu > 0$, y donde ϕ es la CDF de la distribución normal estandar. Ya que $Y = Y_1 + Y_2$ donde $Y_1 \sim N(\mu, \sigma^2)$ y $Y_2 \sim EX(\nu)$ siendo independientes, la media de Y esta dada por $E(Y) = \mu + \nu$ y la varianza por $Var(Y) = \mu^2 + \sigma^2$. Esta distribución también se ha denominado la distribución normal retrasada, Johnson et al. (1994), p172.

Para las distribuciones nombradas anteriormente, se dará a conocer las funciones de enlace dadas por Stasinopoulos *et al* (2008).

Tabla 2-1: Algunas distribuciones continuas para los GAMLSS (junto con sus funciones de enlace).

Distribución	Nombre en R	μ	σ	ν	τ
Exponencial Gaussiana	exGAUS()	Identidad	Log	Log	-
Skew Normal Tipo 2	SN2()	Identidad	Log	Identidad	-
Gumbel Inversa	RG	Identidad	Log	-	-

2.4.2. Selección del modelo

Una vez ajustado el modelo debemos hacer una adecuada selección, una de las funciones de R para esto es la siguiente:

stepGAICAll.A(): La estrategia utilizada por esta función para una distribución dada es ¹³:

- Utilizar el criterio GAIC hacia adelante para seleccionar un modelo apropiado para μ , fijando δ , ν y τ .
- Dado el modelo para μ obtenido en 1 y para ν y τ fijas, úsese un proceso de selección hacia adelante para escoger un modelo apropiado para δ .

¹³Peña, M. J. P. Modelos Aditivos Xeneralizados para Localización, Escala e Forma (GAMLSS).

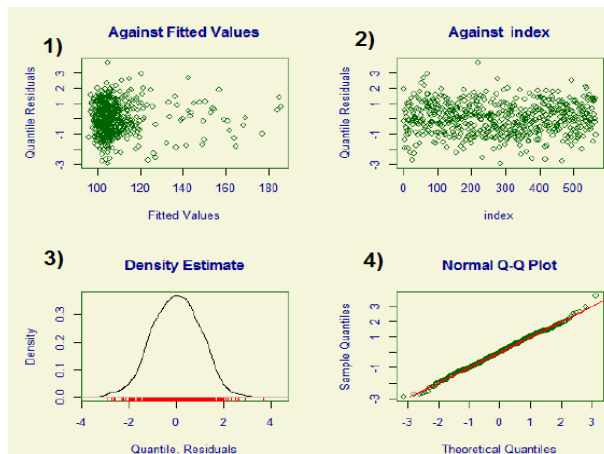
- Dados los modelos para μ y δ obtenidos en 1 y 2 respectivamente y con τ fija, utilizamos un proceso de selección hacia adelante para escoger un modelo apropiado para ν .
- Dados los modelos para μ , δ y ν obtenidos en 1, 2 y 3 respectivamente, utilizamos un proceso de selección hacia adelante para escoger un modelo apropiado para τ .
- Dados los modelos para μ , δ y τ obtenidos en 1, 2 y 4 respectivamente, utilizamos un proceso de selección hacia atrás para escoger un modelo apropiado para ν .
- Dados los modelos para μ , ν y τ obtenidos en 1, 5 e 4 respectivamente, utilizamos un proceso de selección hacia atrás para escoger un modelo apropiado para δ .
- Dados los modelos para δ , ν y τ obtenidos en 6, 5 y 4 respectivamente, usamos un proceso de selección hacia atrás para escoger un modelo apropiado para μ y finalizamos el proceso.

2.4.3. Diagnóstico del modelo

En la figura **2-2**, por medio del comando `plot()` aplicado a la variable respuesta se obtienen cuatro gráficos, los cuales son ¹⁴:

Residuos frente a los valores ajustados, residuos frente alguna variable respuesta del modelo, estimación de kernel y el QQ-Plot.

¹⁴Peña, M. J. P. Modelos Aditivos Generalizados para Localización, Escala e Forma (GAMLSS).

Figura 2-2: Gráficos salida plot()

Fuente: Peña, M. (2015-2015), p 26

La consola R da el valor de los parámetros, los cuales deben ser ajustados a una distribución normal para poder considerar el modelo acertado figura 2-3.

Figura 2-3: Parámetros salida plot()

```
*****
Summary of the Quantile Residuals
      mean   = -6.013267e-05
    variance = 1.000188
  coef. of skewness = 0.009045191
    coef. of kurtosis = 3.279199
Filliben correlation coefficient = 0.9986419
*****
```

Fuente: Peña, M. (2015-2015), p 26

La función `wp ()` (gráfico de gusano) permite identificar el ajuste de los residuales cuantiles aleatorizados a los valores esperados, determinando la adecuada selección de la distribución para la variable respuesta y una descripción de los parámetros 2-2. Donde se considera ¹⁵ :

¹⁵Peña, M. J. P. Modelos Aditivos Xeneralizados para Localización, Escala e Forma (GAMLSS).

Tabla 2-2: Formas de la función $Wp()$.

Forma de wp (worm plot)	Residuos	Variable respuesta
Datos por encima de 0	Media alta	Parámetro para localización bajo
Datos por debajo de 0	Media baja	Parámetro para localización alto
Pendiente positiva	Varianza alta	Parámetro de escala bajo
Pendiente negativa	Varianza baja	Parámetro de escala alto
Forma de U	Asimetría positiva	Asimetría baja
Forma de U inversa	Asimetría negativa	Asimetría alta
Forma de S estirada ala izquierda	Leptocúrtica	Curtosis baja
Forma de S estirada ala iz. volteada	Platicúrtica	Curtosis alta

Fuente: Peña, M. (2015-2015), p 26

La función $Q.stats()$ permite identificar los residuos que no pertenecen a una distribución normal, dando así a conocer los valores atípicos de las variables explicativas continuas del modelo. Teniendo dos salidas, una numérica donde se deben detectar valores mayores a 1,96 y una gráfica de círculos donde el tamaño de estos determinan su valor, si los valores exceden 1,96 aparecerá un recuadro en estos.

2.5 REGRESIÓN LOGÍSTICA ORDINAL (RLO)

En el caso de variables respuesta a las que se les pueda asignar la distribución binomial o multinomial se utilizan los modelos de regresión logística. Las interpretaciones de los coeficientes de β en un modelo de regresión logística se dan por medio de: riesgo relativo, odds y la razón de odds, donde los parámetros, β_1 se estiman de acuerdo al número de individuos existentes en la información del investigador.

Siguiendo los planteamiento de Agresti (2011), para el caso de las variables respuesta con más de dos categorías implicando un orden, la regresión logística ordinal (RLO) es la adecuada, permitiendo la relación de la variable respuesta con las variables explicativas. Así, el desempeño en Comunicación Escrita e Inglés se modelara por medio de la RLO.

Para McCullagh (1980), la función de enlace logit es la acertada para manejar datos ordinales cuya distribución de frecuencia es uniforme en las categorías de la variable

respuesta. La expresión de esta función para la RLO es la siguiente:

Con Y , la variable respuesta y $X\beta$ la matriz de las variables explicativas

$$\log(O_i) = \alpha i + \beta X \quad (2-10)$$

donde

$$\log\left(\frac{P(Y \leq j|x)}{1 - P(Y \leq j|x)}\right) = \alpha i + \beta X \quad (2-11)$$

Una forma de utilizar el orden en categorías logits de probabilidades acumulativas, es ¹⁶:

$$P(Y \leq j|x) = \pi_1(x) + \dots + \pi_j(x) \quad (2-12)$$

Los logits acumulados son:

$$\text{logit}[P(Y \leq j|x)] = \log \frac{P(Y \leq j|x)}{1 - P(Y \leq j|x)} \quad (2-13)$$

$$= \log \frac{\pi_1(x) + \dots + \pi_j(x)}{\pi_{j+1}(x) + \dots + \pi_J(x)} \quad (2-14)$$

con

$$j = 1, \dots, J - 1. \quad (2-15)$$

Cada logit acumulativo utiliza todas las categorías de respuesta J (categoría de respuesta para j). Es relevante tener en cuenta que si la variable de estudio posee j categorías, las ecuaciones que se pueden generar serán una categoría menor ($j-1$).

Para la estimación de los coeficientes y de sus errores estandar se emplea la estimación por máxima verosimilitud, expresada por:

$$\prod_{i=1}^n \left[\prod_{j=1}^J \pi_j(x_i)^{y_{ij}} \right] = \prod_{i=1}^n \left[\prod_{j=1}^J (P(Y \leq j|x_i) - P(Y \leq j-1|x_i))^{y_{ij}} \right] \quad (2-16)$$

$$= \prod_{i=1}^n \left[\prod_{j=1}^J \left(\frac{\exp(\alpha_i + \beta' X_i)}{1 + \exp(\alpha_i + \beta' X_i)} - \frac{\exp(\alpha_{j-1} + \beta' X_i)}{1 + \exp(\alpha_{j-1} + \beta' X_i)} \right)^{y_{ij}} \right] \quad (2-17)$$

¹⁶Agresti, A., & Kateri, M. (2011). Categorical data analysis. In International encyclopedia of statistical science (pp. 206-208). Springer Berlin Heidelberg.

Vista como una función de $(\{\alpha_i\}, \beta)$, donde se utiliza el algoritmo de Fisher scoring para obtener las estimaciones de Maxima verosimilitud ¹⁷. McCullang (1980) y Walker Y Duncan (1967).

Un odds de probabilidades acumulativas se denomina una razón de probabilidades acumulativa. Esta razón de probabilidades acumulativas se resume a un cociente entre la probabilidad que ocurra un suceso frente a la probabilidad que no ocurra. Expresada por ¹⁸:

$$\text{logit} [P(Y \leq j|x_1)] - \text{logit} [P(Y \leq j|x_2)] \quad (2-18)$$

$$= \log \frac{P(Y \leq j|x_1)/P(Y \leq j|x_2)}{P(Y \leq j|x_2)/P(Y \leq j|x_1)} \quad (2-19)$$

$$= \beta'(x_1 - x_2) \quad (2-20)$$

La razon de probabilidades acumulativa se utiliza para identificar el efecto de las variables explicativas en la variable respuesta. Según Agresti (1990), “si éste es igual a uno, indica que la variable respuesta no tiene efecto. Si es menor que uno, lo cual sucede cuando el coeficiente de la variable explicativa es negativo, indica que, si las otras variables regresoras permanecen constantes, los cambios en la variable explicativa analizada incrementan la probabilidad de obtener categorías de mayor valor en la variable objeto de estudio (Hosmer y Lemeshow, 2000)”. Valores de odds mayores que uno muestran que las variaciones en la variable independiente disminuyen la probabilidad de obtener categorías de mayor valor de la dependiente¹⁹.

La significancia individual de cada variable independiente, generalmente, se analiza a través de la prueba de Wald, la cual se basa en la significancia del coeficiente estimado (estimación entre el error estandar) para cada variable ²⁰. El estadístico empleado es el

¹⁷Agresti, A., & Kateri, M. (2011). Categorical data analysis. In International encyclopedia of statistical science (pp. 206-208). Springer Berlin Heidelberg.

¹⁸Agresti, A., & Kateri, M. (2011). Categorical data analysis. In International encyclopedia of statistical science (pp. 206-208). Springer Berlin Heidelberg.

¹⁹Gonzalez, J. J. G. (2017, p 1- 31.), ‘Aplicación de un modelo de regresión logistica con respuesta politómica ordinal en el análisis del desempeño académico en matemáticas’.

²⁰Agresti, A., & Kateri, M. (2011). Categorical data analysis. In International encyclopedia of statistical science (pp. 206-208). Springer Berlin Heidelberg.

siguiente:

$$Z_{Wald} = \left(\frac{b_j}{ES(b_j)} \right) \quad (2-21)$$

3 DISEÑO METODOLÓGICO

Este estudio se enfoca en determinar los factores influyentes en el puntaje de las pruebas Saber Pro presentadas por estudiantes del programa Licenciatura en Matemáticas y Estadística desde el segundo semestre del 2011 al 2015, teniendo como fuente principal de información las bases de datos del ICFES. El estudio se clasifica como una investigación descriptiva, ya que se interpretó los datos brindados por el ICFES en el programa de Licenciatura en Matemáticas y Estadística, buscando determinar los factores demográficos, socioeconómicos y académicos que inciden en el resultado del componente genérico de las pruebas Saber Pro. El estudio presentó un enfoque cuantitativo, ya que para la interpretación se utilizaron herramientas estadísticas.

3.1 POBLACIÓN Y TIPO DE DISEÑO

La población objetivo son los puntajes en el componente genérico de la prueba Saber Pro presentadas por los estudiantes del programa de Licenciatura en Matemáticas y Estadística pertenecientes a la Universidad Pedagógica y Tecnológica de Colombia Facultad Seccional Duitama, que presentaron el examen entre el segundo semestre del 2011 y el 2015.

Este estudio tiene un tipo de diseño bibliográfico, ya que se trabajó con las bases de datos del ICFES, es decir, se utilizó datos secundarios que han sido recogidos y procesados. Nos centramos en los datos correspondientes a las pruebas Saber Pro, las cuales fueron facilitadas en los seminarios orientados por el ICFES, estos datos constan de una serie de archivos, los cuales son: Guías de orientación de los años donde se presentaron, datos a nivel Colombia de todas las instituciones de educación superior y técnica, acompañado por diccionarios donde se especifica cada variable y sus posibles valores ¹.

3.2 MÉTODOS DE ANÁLISIS Y PROCESAMIENTO DE DATOS

Para este estudio se utilizaron técnicas de tipo descriptivo, pruebas de diferencia de medias, pruebas de correlación, tablas de contingencia, pruebas de bondad de ajuste. También se emplearon técnicas de modelamiento estadístico como: GAMLSS y regresión logística ordinal. Para el procesamiento de los datos se utilizó el software libre R como procesador estadístico versión 3.3.0, (como uno de los lenguajes más utilizados en investigación por la comunidad estadística), se trabajó con los siguientes paquetes: gamlss, MASS, gamlss.dist, gamlss.data, fitdistrplus y Rcmdr.

¹Guías, I. C. F. E. S. (2011). Orientaciones para el examen de Estado de calidad de la educación superior SABER PRO (ECAES). Bogotá, DC, septiembre, ICFES. www.icfes.gov.co. ICFES. Programas por año Saber Pro. Bogotá D.C. ICFES. Guías de orientación, 2004 - 2009, Bogotá D.C, ICFES. Diccionario de variables Saber Pro Pruebas Genéricas Periodos 2012-1 a 2013-2 (Universitarios), Bogotá D.C, ICFES. Programas por año Saber Pro. Bogotá D.C, ICFES. Bases de datos, 2006- 2015, Bogotá D.C, ICFES. (2011). Orientaciones para el examen de Estado de calidad de la educación superior Saber Pro (ECAES). Bogotá D.C, ICFES. Diccionario de datos Saber Pro (ECAES). Bogotá D.C, Diccionario de variables Saber Pro Pruebas Genéricas Periodos 2012-1 a 2013-2 (Universitarios), ICFES. Programas por año Saber Pro. Bogotá D.C.

3.3 VARIABLES DE ESTUDIO

Las variables de estudio se dividieron en tres grupos: Factores Académicos, Factor Demográficos y Factor Socioeconómicos.

Factor Académico

A esta categoría pertenecen aquellas variables asociadas al rendimiento académico previo y al esfuerzo mostrado por el estudiante, así como al currículo que sigue el mismo. Así, abarcan aquellos factores educativos que rodean directamente la intervención pedagógica o docente.

Factor Demográfico

A esta categoría pertenecen aquellas variables que se refieran a características de la población tales como: sexo, edad, lugar de residencia, estado civil, cabeza de familia, situación del hogar entre otras.

Factor Socioeconómico

A esta categoría pertenecen aquellas variables asociadas al aspectos económicos tales como: estrato socioeconómico, forma de pago de la matrícula, nivel de sisben, estrato de residencia, ocupación de los padres, celular, internet, servicio de televisión, computador, lavadora, salario mensual familiar, número de dormitorios y personas a cargo, horas de trabajo, trabaja entre otras.

A continuación se presenta de cada factor sus variables, tipo, escala de medición y la descripción.

Figura 3-1: Variables de estudio Factor Académico.

NOMBRE DE LA VARIABLE	TIPO DE VARIABLE/ESCALA DE MEDICIÓN	CATEGORÍA/RANGO /EJEMPLO	DESCRIPCIÓN
Porcentaje créditos Codificada ESTU_PJE_CREDITOS	CUALITATIVA NOMINAL	valores posibles: 5	Porcentaje de créditos necesarios para optar al título que ha cursado y aprobado hasta el momento
		0	No sigue sistema de créditos
		1	Menos del 75%
		2	Entre el 76% y el 80%
		3	Entre el 81% y el 90%
		4	El 90% o más
Tipo de colegio: Codificada NST_ORIGEN_COD	CUALITATIVA NOMINAL	valores posibles: 6	Origen de la institución
		1	Oficial nacional
		2	Oficial Departamental
		3	Oficial Municipal
		7	No oficial – Fundación
		8	No oficial – Corporación
		9	Régimen especial
Tomar Curso refuerzo Codificada :ESTU_ACTIVIDADNOCURSO	CUALITATIVA NOMINAL	valores posibles: 2	Si no tomó ningún curso
		0	ninguna actividad de preparación para el examen
		1	Repasó por cuenta propia
Tomar curso IES Codificada: ESTU_CURSODOCENTESIES	CUALITATIVA NOMINAL	valores posibles: 4	Tomó curso en su IES con docentes de la institución
		0	No tomó curso
		1	Menos de 20 horas
		2	Entre 20 y 30 horas
		3	Más de 30 horas
Tomar curso por externo Codificada :ESTU_CURSOIESEXTERNO	CUALITATIVA NOMINAL	valores posibles: 4	Tomó curso organizado por la institución con apoyo externo
		0	No tomó curso
		1	Menos de 20 horas
		2	Entre 20 y 30 horas
		3	Más de 30 horas
Tomar curso instituto: Codificada: ESTU_INSTITUTOPREPARACION	CUALITATIVA NOMINAL	valores posibles: 4	Tomó curso en un instituto de preparación de exámenes
		0	No tomó curso
		1	Menos de 20 horas
		2	Entre 20 y 30 horas
		3	Más de 30 horas
Tomar simulacro ICFES Codificada :ESTU_CURSOPREGICFES	CUALITATIVA NOMINAL Dicotómica	valores posibles: 2	Simulacro con preguntas tipo ICFES
		0	No
		1	Si
Tomar refuerzos: Codificada :ESTU_REFUERZOREAS	CUALITATIVA NOMINAL Dicotómica	valores posibles: 2	Clases de refuerzo en las diferentes áreas
		0	No
		1	Si
Refuerzo competencias genéricas Codificada :ESTU_REFUERZOGSA	CUALITATIVA NOMINAL Dicotómica	valores posibles: 2	Refuerzo en desarrollo de competencias genéricas
		0	No
		1	Si
Tipo de bachillerato Codificada :ESTU_TITULO_BTO	CUALITATIVA ORDINAL	valores posibles: 4	Tipo de bachillerato del que se graduó el evaluado
		A	Académico
		N	Técnico
		T	Normalista Superior
			Desconocido

Fuente: Autor.

Continuación figura 3-1

Figura 3-2: Variables de estudio Factor Académico.

NOMBRE DE LA VARIABLE	TIPO DE VARIABLE/E ESCALA DE MEDICIÓN	CATEGORÍA/RANGO /EJEMPLO	DESCRIPCIÓN
Semestre de presentación al examen: Codificada: ESTU_EXAM_SEMESTRE_PRESTACION	CUALITATIVA ORDINAL	valores posibles: 2	Semestre de presentación al examen
		1	Primer semestre
		2	Segundo semestre
Puntaje Comunicación escrita Codificada: MOD_COMUNICA_ESCRITA_PUNT	CUANTITATIVA CONTINUA	rango: [6.8, 13.1]	Puntaje Comunicación escrita
Desempeño Comunicación escrita Codificada : MOD_COMUNICA_ESCRITA_DESEM	CUANTITATIVA CONTINUA	rango: [1,8]	Desempeño Comunicación escrita
Puntaje Inglés Codificada: :MOD_INGLES_PUNT	CUANTITATIVA CONTINUA	rango: [0,15]	Puntaje Inglés
Desempeño Inglés Codificada: MOD_INGLES_DESEM	CUALITATIVA ORDINAL	valores posibles: 5	Desempeño Inglés
		A-	A-
		A1	A1
		A2	A2
		B+	B+
		B1	B1
Puntaje Lectura crítica Codificada: :MOD_LECTURA_CRITICA	CUANTITATIVA CONTINUA	rango: [0,14.9]	Puntaje Lectura crítica
Puntaje Razonamiento Cuantitativo	CUANTITATIVA CONTINUA	rango: [0,16.8]	Puntaje Razonamiento cuantitativo
Máximo nivel educativo padre/madre Codificada: FAMI_COD_EDUCA_PADRE FAMI_COD_EDUCA_MADRE	CUALITATIVA ORDINAL	Valores posibles:19	Máximo nivel educativo alcanzado por el padre
		0	Ninguno
		1	No tuvo escuela
		2	Preescolar
		3	Básica primaria
		4	Básica secundaria
		5	Media vocacional
		6	Tecnológico o técnico
		7	Universitario
		8	Postgrado
		9	Primaria incompleta
		10	Primaria completa
		11	Secundaria (bachillerato) incompleta
		12	Secundaria (bachillerato) completa
		13	Educación técnica o tecnológica incompleta
		14	Educación técnica o tecnológica completa
		15	Educación profesional incompleta
		16	Educación profesional completa
		17	Postgrado
		99	No sabe

Fuente: Autor.

Figura 3-3: Variables de estudio Factor Demográfico.

CLASIFICACIÓN DE VARIABLES	TIPO DE VARIABLE/ESCALA DE MEDICIÓN	CATEGORÍA/RANGO/EJEMPLO	DESCRIPCIÓN
Edad. Codificada: ESTU_NACIMIENTO_ANNO	CUANTITATIVA CONTINUA	Rango: [1934,2002]	Año de nacimiento del inscrito
Estado civil Codificada: ESTU_ESTADO_CIVIL	CUALITATIVA NOMINAL	Valores posibles: 5	Estado civil del estudiante
		1	Soltero
		2	Casado
		3	Viudo
		4	Separado
		5	Unión libre
Sexo Codificada: ESTU_GENERO	CUALITATIVA NOMINAL	Valores posibles: 2	Sexo del estudiante
		M	Masculino
		F	Femenino
Número de personas hogar Codificada: FAMI_NUM_PERS_GRUP_FAM	CUANTITATIVA DISCRETA DE RAZÓN	rango: [1,12]	Número de personas que conforman el hogar
Número de personas a cargo Codificada: FAMI_NUM_PERS_CARGO	CUANTITATIVA DISCRETA DE RAZÓN	rango: [0,5]	Número de personas de las que usted se encuentra a cargo
Cabeza de familia Codificada: ESTU_SN_CABEZA_FMLIA	CUALITATIVA NOMINAL DICOTOMICA	valores posibles: 2	Es cabeza de familia
		0	No
		1	Si
Situación de su hogar Codificada: ESTU_HOGAR_ACTUAL	CUALITATIVA NOMINAL	valores posibles: 2	Situación de su hogar actual
		1	Es habitual o permanente
		2	ES temporal por razones de estudio u otra razón

Fuente: Autor.

Figura 3-4: Variables de estudio Factor Socioeconómico.

CLASIFICACIÓN DE VARIABLES	TIPO DE VARIABLE/ ESCALA DE MEDICIÓN	CATEGORÍA/RANGO /EJEMPLO	DESCRIPCIÓN
Estrato de la residencia de acuerdo a la factura de energía Codificada: ESTU_ESTRATO	CUALITATIVA ORDINAL	valores posibles: 7	Estrato socioeconómico de la residencia del estudiante según factura de energía
		1	Estrato 1
		2	Estrato 2
		3	Estrato 3
		4	Estrato 4
		5	Estrato 5
		6	Estrato 6
		8	Vive en una zona rural donde no hay estratificación económica
Nivel de SISBEN Codificada: FAMI_NIVEL_SISBEN	CUALITATIVA ORDINAL	valores posibles: 5	Nivel de SISBEN en que está clasificada la familia
		1	Nivel 1
		2	Nivel 2
		3	Nivel 3
		4	Clasificado en otro nivel
Celular Codificada: ECON_SN_CELULAR	CUALITATIVA NOMINAL DICOTÓMICA	valores posibles: 2	El hogar cuenta con celular
		0	No
Conexión internet Codificada: ECON_SN_INTERNET	CUALITATIVA NOMINAL DICOTÓMICA	valores posibles: 2	El hogar cuenta con conexión a internet
		0	No
Televisión Codificada: ECON_SN_SERVICIO_TV	CUALITATIVA NOMINAL DICOTÓMICA	valores posibles: 2	El hogar cuenta con Servicio cerrado de televisión
		0	No
Teléfono fijo Codificada: ECON_SN_TELEFONIA	CUALITATIVA NOMINAL DICOTÓMICA	valores posibles: 2	El hogar cuenta con servicio de teléfono fijo
		0	No
Lavadora Codificada: ECON_SN_LAVADORA	CUALITATIVA NOMINAL DICOTÓMICA	valores posibles: 2	El hogar cuenta con lavadora
		0	No
Computador Codificada: ECON_SN_COMPUTADOR	CUALITATIVA NOMINAL DICOTÓMICA	valores posibles: 2	Tiene computador en su hogar
		0	No
Número de dormitorios Codificada: ESTU_DORMITORIOS	CUANTITATIVA DISCRETA DE RAZON	1-10	Número de dormitorios de la residencia
Salario mensual Codificada: FAMI_ING_FMLIAR_MENSUAL	CUANTITATIVA CONTINUA DE RAZON	valores posibles: 7	Ingresos mensuales representado en salarios mínimos mensuales
		1	Menos de 1 SM
		2	Entre 1 y Menos de 2 SM
		3	Entre 2 y Menos de 3 SM
		4	Entre 3 y Menos de 5 SM
		5	Entre 5 y Menos de 7 SM
		6	Entre 7 y Menos de 10 SM
		7	10 o más SM
Tiene trabajo Codificada: ESTU_TRABAJA	CUALITATIVA NOMINAL	valores posibles: 4	Trabaja actualmente
		0	No
		1	Si, con remuneración en dinero o en especie
		3	Si, como ayudante sin remuneración
		4	Si, por ser práctica obligatoria del plan de estudios
		5	Si, para adquirir experiencia y/o recursos para sus gastos personales

Fuente: Autor.

Continuación figura 3-4

Figura 3-5: Variables de estudio Factor Socioeconómico.

CLASIFICACIÓN DE VARIABLES	TIPO DE VARIABLE/ESCALA DE MEDICIÓN	CATEGORÍA/RANGO /EJEMPLO	DESCRIPCIÓN
Recursos propios Codificada: ESTU_SN_MATRICULA_PROPIO	CUALITATIVA NOMINAL DICOTÓMICA	valores posibles: 2	Los recursos con los que canceló la matrícula el año pasado fueron propios
		0	No
		1	Si
Recursos padres Codificada: ESTU_SN_MATRICULA_PADRES	CUALITATIVA NOMINAL DICOTÓMICA	valores posibles: 2	Los recursos con los que canceló la matrícula el año pasado provinieron de padres
		0	No
		1	Si
Crédito Codificada: ESTU_SN_MATRICULA_CREDITO	CUALITATIVA NOMINAL DICOTÓMICA	valores posibles: 2	Los recursos con los que canceló la matrícula el año pasado provinieron de crédito
		0	No
		1	Si
Beca Codificada: ESTU_SN_MATRICULA_BECA	CUALITATIVA NOMINAL DICOTÓMICA	valores posibles: 2	Los recursos con los que canceló la matrícula el año pasado provinieron de beca
		0	No
		1	Si
Valor matrícula Codificada: INST_VLR_MATRICULA_ANT	CUALITATIVA NOMINAL	valores posibles: 6	Valor anual de la matrícula del año anterior
		0	No pagó matrícula
		1	Menos de 500 mil
		2	Entre 500 mil y menos de 1 millón
		3	Entre 1 millón y 3 millones
		4	Entre 3 millones y 5 millones
		5	Más de 5 millones
Horas trabajadas semanal Codificada: ESTU_HORAS_TRABAJO	CUANTITATIVA DISCRETA	8-20	Número de horas que trabaja a la semana
Ocupación del padre Codificada: FAMI_COD_OCUP_PADRE /FAMI_COD_OCUP_MADR E	CUALITATIVA NOMINAL	valores posibles: 27	Ocupación del padre
		1	Empresarios
		2	Administradores o gerentes
		3	Profesionales independientes
		4	Profesionales empleados
		5	Trabajadores independientes
		6	Trabajadores empleados
		7	Rentistas
		8	Obreros
		9	Jubilados
		10	Hogar
		11	Estudiantes
		12	No devengan ingreso o buscan trabajo
		13	Empresario
		14	Pequeño empresario
		15	Empleado con cargo como director o gerente general
		16	Empleado de nivel directivo
		17	Empleado de nivel técnico o profesional
		18	Empleado de nivel auxiliar o administrativo
		19	Empleado obrero u operario
		20	Profesional Independiente
		21	Trabajador por cuenta propia
		22	Hogar
		23	Pensionado
		24	Rentista
		25	Estudiante
		26	Otra actividad u ocupación
		99	No sabe

Fuente: Autor.

4 DESCRIPCIÓN DE LOS DATOS

A continuación se describe el desempeño de los estudiantes del Programa en las Pruebas Saber Pro. En primer lugar, se describen los resultados, entre los años 2011 (segundo semestre) y el año 2015, discriminados por competencias. Posteriormente se presentan los resultados del año 2016. Los cuales fueron tomados de la pagina <http://www.icfes.gov.co/instituciones-educativas-y-secretarias/saber-pro/resultados-agregados> .

4.1 RESULTADOS COMPETENCIAS GENÉRICAS

Las competencias genéricas son las que se consideran debe desarrollar todo estudiante, independiente de su formación profesional, se evalúa Inglés, Razonamiento Cuantitativo, Lectura Crítica, Comunicación Escrita y Competencias Ciudadanas. Los resultados de cada módulo correspondiente a las Licenciaturas en el País y a las de la UPTC se presentan para cada competencia, n , el número de estudiantes que presentaron la prueba, \bar{x} , puntaje promedio obtenido, DE , desviación estándar, CA , coeficiente de asimetría, CC , coeficiente de curtosis, CV , coeficiente de variación, C_1 , C_2 , C_3 representan los cuartiles uno, dos y tres respectivamente y NA se refiere al número de datos faltantes. Los resultados se presentan en tablas las cuales incluyen los resultados correspondientes al Programa.

Tabla 4-1: Resultados módulo Inglés

Año	n	\bar{x}	CA	CC	CV	C_1	C_2	C_3	NA
2011	12	9.9	0.3	-0.4	0.07	9.5	9.8	10.2	0
2012	18	9.7	1.1	0.7	0.06	9.4	9.6	9.8	0
2013	29	10	0.7	-0.4	0.08	9.5	9.8	10.3	0
2014	28	9.9	0.8	0.7	0.07	9.5	9.9	10.1	0
2015	9	9.8	-0.6	-0.01	0.06	9.6	10.1	10.2	0

Tabla 4-2: Resultados módulo Inglés

Año	Promedio Licenciaturas (\bar{x} (DE))	Promedio Licenciaturas UPTC-Tunja, (\bar{x} (DE))
2011	NA	NA
2012	9.87(1.42)	10.02(1.21)
2013	9.91(1.47)	9.98(1.21)
2014	9.90(1.43)	10.11(1.31)
2015	10.08(1.53)	10.01(1.18)

En la Tablas 4-1 y 4-2 se presentan los resultados obtenidos en INGLÉS. Se observa que, para cada uno de los años que han presentado la prueba, los estudiantes de Licenciatura en Matemáticas y Estadística son muy homogéneos respecto a los puntajes (la variabilidad relativa, coeficientes de variación, son inferiores al 8%). Los coeficientes de curtosis indican que no hay estudiantes que hayan obtenido puntajes muy diferentes a

los del grupo con los que presentaron la prueba. Sin considerar el año de la presentación de la prueba y teniendo en cuenta que el promedio es de 9.88, aproximadamente el 68 % de los estudiantes obtuvieron resultados entre 9.19 y 10.57 puntos, lo cual los ubica en la categoría regular de la escala ¹.

Los resultados anteriores se pueden complementar con los resultados cualitativos en INGLÉS, Tabla 4-3. Sin tener en cuenta el año de presentación de la prueba, se tiene que la mayor proporción de estudiantes (40 %) se ubican en el nivel de desempeño A1, indicando que son capaces de comprender y utilizar expresiones cotidianas de uso muy frecuente así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato. De otro lado se tiene que el 17 % de los estudiantes se ubican en el nivel A2 y el mismo porcentaje en el nivel más alto de desempeño, B1.

Tabla 4-3: Resultados de Desempeño en Inglés

Año	A-	A1	A2	B1	n	NA
2011	25 %	33 %	33 %	8 %	12	0
2012	39 %	44 %	6 %	11 %	18	0
2013	24 %	35 %	17 %	24 %	29	0
2014	25 %	50 %	7 %	18 %	28	0
2015	22 %	22 %	44 %	11 %	9	0

Los resultados de RAZONAMIENTO CUANTITATIVO se presentan en la Tablas 4-4 y 4-5. Se observa que, para cada uno de los años que han presentado la prueba, los estudiantes de Licenciatura en Matemáticas y Estadística son muy homogéneos (la variabilidad relativa, coeficientes de variación, son inferiores al 9 %). Los coeficientes de curtosis indican que no hay estudiantes que hayan obtenido puntajes muy diferentes a los del grupo con los que presentaron la prueba. Sin considerar el año de la presentación de la prueba y teniendo en cuenta que el promedio es de 10.50 y la desviación estándar es de 0.90 puntos, aproximadamente el 68 % de los estudiantes obtuvieron resultados entre 9.61 y 11.40 puntos, lo cual los ubica en la categorías de regular, bueno y muy bueno en la escala.

¹Sánchez, Ivan de Jesus. Guía de interpretación de resultados de las pruebas Saber Pro. 2013

Tabla 4-4: Resultados Razonamiento Cuantitativo

Año	n	\bar{x}	CA	CC	CV	C_1	C_2	C_3	NA
2011	12	10.3	-0.2	-1.2	0.07	9.9	10.3	11.0	0
2012	18	10.3	-0.1	-1.0	0.06	9.9	10.3	10.3	0
2013	29	10.7	0.1	-0.7	0.09	9.9	10.5	11.5	0
2014	28	10.6	0.8	-0.1	0.09	9.8	10.4	11.3	0
2015	9	10.8	0.4	-0.6	0.07	9.9	10.5	10.8	0

Tabla 4-5: Resultados Razonamiento Cuantitativo

Año	Promedio Licenciaturas, (\bar{x} (DE))	Promedio Licenciaturas UPTC-Tunja, (\bar{x} (DE))
2011	NA	NA
2012	9.48(0.82)	9.65(0.78)
2013	9.52(0.84)	9.61(0.77)
2014	9.47(0.86)	9.69(0.80)
2015	9.54(0.97)	9.71(0.85)

Los resultados de COMUNICACIÓN ESCRITA se presentan en la Tablas 4-6 y 4-7. Se observa que, para cada uno de los años que han presentado la prueba, los estudiantes de Licenciatura en Matemáticas y Estadística son homogéneos (la variabilidad relativa, coeficientes de variación, son inferiores al 11 %). Los coeficientes de curtosis indican que no hay estudiantes que hayan obtenido puntajes muy diferentes a los del grupo con los que presentaron la prueba. Sin considerar el año de la presentación de la prueba y teniendo en cuenta que el promedio es de 9.91 y la desviación estándar es de 0.89 puntos, aproximadamente el 68 % de los estudiantes obtuvieron resultados entre 9.02 y 10.80 puntos, lo cual los ubica en la categoría de regular y buena en la escala.

Tabla 4-6: Resultados comunicación escrita

Año	n	\bar{x}	CA	CC	C	C_1	C_2	C_3	NA
2011	11	10.1	-0.3	-1.4	0.08	9.4	10.5	10.85	1
2012	18	9.8	-1.0	0.7	0.11	9.3	10.4	10.7	0
2013	28	10.0	-0.6	0.6	0.08	9.7	10.2	10.5	1
2014	28	9.6	0.1	-0.7	0.07	9.2	9.4	10.2	0
2015	9	10.1	-1.3	0.3	0.09	10.3	10.4	10.8	0

Tabla 4-7: Resultados comunicación escrita

Año	Promedio Licenciaturas, (\bar{x} (DE))	Promedio Licenciaturas UPTC-Tunja, (\bar{x} (DE))
2011	NA	NA
2012	10.13(1.22)	10.34(1.06)
2013	9.95(1.04)	10.17(0.92)
2014	9.88(1.04)	10.19(0.94)
2015	9.90(1.07)	10.13(0.86)

Los resultados anteriores se pueden complementar con los resultados cualitativos en comunicación escrita, Tabla 4-8. Sin tener en cuenta el año de presentación de la prueba, se tiene que la mayor proporción de estudiantes (44 %) se ubican en el Nivel 5 de desempeño, indicando que encadenan o relacionan efectivamente las ideas, dándole continuidad al escrito: incorpora información nueva vinculándola con la anterior, presentando la información en un orden cronológico, partiendo de temas generales para desglosar temas específicos, entre otras maneras de interrelacionar contenidos. De otro lado se tiene que el 29 % de los estudiantes se ubican en el Nivel 4 y muy pocos (4 %) en el Nivel 6. No se tienen estudiantes en el Nivel 7 u 8.

Tabla 4-8: Desempeño en comunicación escrita

Año	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	n	NA
2011	0 %	0 %	27 %	18 %	45 %	9 %	11	1
2012	6 %	6 %	16 %	16 %	56 %	0 %	18	0
2013	0 %	7 %	36 %	7 %	39 %	11 %	28	1
2014	0 %	0 %	29 %	42 %	29 %	0 %	28	0
2015	0 %	0 %	22 %	0 %	78 %	0 %	9	0

Los resultados de LECTURA CRÍTICA se presentan en la Tablas 4-9 y 4-10. Se observa que, para cada uno de los años que han presentado la prueba, los estudiantes de Licenciatura en Matemáticas y Estadística son muy homogéneos (la variabilidad relativa, coeficientes de variación, son inferiores al 10 %); excepto en el año 2015, donde se observa una gran heterogeneidad en los puntajes. Los coeficientes de asimetría indican que no hay estudiantes que hayan obtenido puntajes muy diferentes a los del grupo con los que presentaron la prueba, excepto en el año 2015 en donde se observa que hubo estudiantes con puntaje superior a los de la mayoría del grupo. Los valores indican que entre el año 2011 y el 2014 el puntaje promedio de los estudiantes ha estado entre 9.8

y 10.3, lo cual los ubica en la categoría de regular en la escala. En el año 2015, el 25 % de los estudiantes obtuvo puntajes superiores a 11.2, lo cual los ubica en la categoría muy bueno de la escala. De hecho el 75 % de los estudiantes obtuvo puntajes superiores a 10.2, lo cual indica su buen desempeño.

Tabla 4-9: Resultados de Lectura Crítica

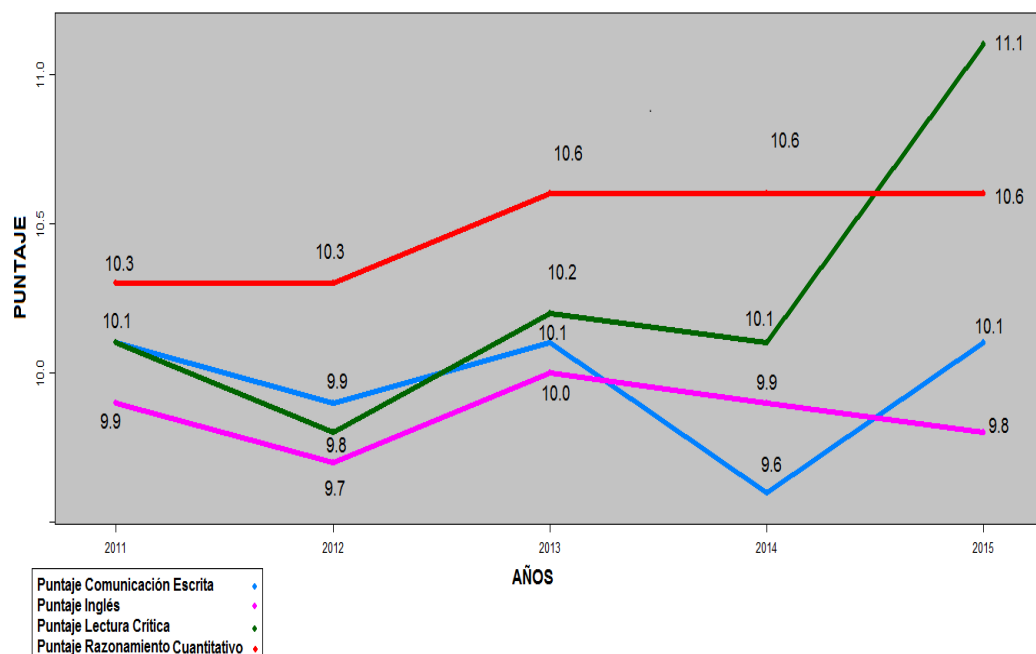
Año	n	\bar{x}	CA	CC	CV	C_1	C_2	C_3
2011	12	10.1	-0.7	0.10	9.3	9.8	10.8	0
2012	18	9.8	0.4	0.08	9.3	9.9	10.1	0
2013	29	10.3	0.03	0.09	9.8	10.5	10.9	0
2014	28	10.1	-0.1	0.08	9.4	10.2	10.9	0
2015	9	11.1	2.0	0.18	10.2	10.5	11.2	0

Tabla 4-10: Resultados de Lectura Crítica

Años	Promedio Licenciaturas, (\bar{x} (DE))	Promedio Licenciaturas UPTC-Tunja, (\bar{x} (DE))
2011	NA	NA
2012	9.65(1.03)	9.79(0.88)
2013	9.81(1.06)	9.89(0.91)
2014	9.71(1.07)	9.94(0.98)
2015	10.19(0.98)	10.26(0.81)

La figura 4-1 muestra el comportamiento de los resultados en el componente genérico desde el ao 2011 hasta el segundo semestre del 2015.

Figura 4-1: Desempeño Competencias Genéricas.



De acuerdo con la figura anterior, el componente de Razonamiento Cuantitativo es en el que mejor les va a los estudiantes de Licenciatura en Matemáticas y Estadística, manteniendo un puntaje constante en los años 2013 a 2015, mientras que Inglés es donde obtienen puntajes más bajos, en el caso de los componentes de Lectura Crítica y Comunicación Escrita se presentó poca variación en los años 2011 a 2013. Comunicación Escrita tuvo el puntaje mas bajo en el año 2014 con relación al periodo, en cambio Lectura Crítica mostró un crecimiento relevante en el año 2015, siendo el desempeño más alto de los estudiantes.

La escala de medición para los resultados de la prueba Saber Pro del año 2016, fue diferente a la de los años anteriores (2011-2015), por tal motivo se realizó una sección aparte, la cual se presentará a continuación:

Para el año 2016, el puntaje global del programa fue de 150 puntos con una desviación estándar de 15 puntos, esto implica que los estudiantes de la Licenciatura fueron muy

homogéneos (coeficiente de variación del 10 %) en la prueba y que al menos el 75 % de ellos obtuvo entre 120 y 180 puntos. La media en las licenciaturas del País fue de 142 puntos con una desviación de 23. Es decir, al menos el 75 % de los estudiantes obtuvo entre 96 y 188 puntos. Lo cual implica que la licenciatura en Matemáticas y Estadística tiene un comportamiento similar a las demás licenciaturas del País. El puntaje global de la Licenciatura también fue similar con el de la Institución y el de la Sede.

4.1.1. Resultados en competencias genéricas año 2016

En competencias genéricas se tiene que en **Comunicación Escrita** el puntaje del Programa (136 ($DE=31$)) estuvo por debajo de los puntajes de la Institución, la Sede y las licenciaturas del País. La mayoría de estudiantes del Programa (52 %) se ubica en nivel de desempeño 2, lo que indica que tan sólo superan las preguntas de menor complejidad del examen y los ubica en un uso aceptable del lenguaje, identificando un planteamiento o posición personal que va más allá de una opinión aislada sobre el tema, sin embargo el texto presentado puede carecer de unidad semántica o presencia de contradicciones, digresiones y repeticiones.

En **Razonamiento Cuantitativo**, el Programa fue superior ($170(DE = 24)$) a las licenciaturas del país y a la Institución, fue similar con los programas de la Sede. La mayoría de estudiantes del Programa (62 %) se ubica en nivel de desempeño 3, lo que indica que muestran un desempeño adecuado en las competencias exigibles para el examen. Este es el nivel esperado que todos o la gran mayoría de los estudiantes deberían alcanzar. Indicando que los estudiantes extraen información implícita contenida en representaciones no usuales asociadas a una misma situación y provenientes de una única fuente de información, argumentan la validez de procedimientos, y resuelven problemas utilizando modelos que combinan procedimientos aritméticos, algebraicos, variacionales y aleatorios. Como se tiene que el 53 % de los estudiantes respondió incorrectamente a la afirmación 1, las acciones de mejoramiento del Programa se deben enfocar en identificar las dificultades a la hora de validar procedimientos y estrategias matemáticas para dar solución a problemas.

En **Lectura Crítica** e **Inglés** el rendimiento del programa fue similar con los de la Institución, la Sede y las licenciaturas del País. En lectura crítica la mayor proporción de estudiantes del Programa (48 %) se ubica en nivel de desempeño 2, lo que indica que tan sólo superan las preguntas de menor complejidad del examen y muestra que los

estudiantes reconocen la macro estructura del texto e identifica la tipología textual, las estrategias discursivas, y reconoce las funciones del lenguaje para comprender el sentido del texto. Como se tiene que el 49 % de los estudiantes respondió incorrectamente a la afirmación 2, las acciones de mejoramiento del Programa se deben enfocar identificar las dificultades a la hora de articular partes de un texto para darle un sentido global.

En **Inglés** la mayor proporción de estudiantes del Programa (43 %) se ubica en nivel de desempeño A1, es decir, los estudiantes tan sólo son capaces de comprender y utilizar expresiones cotidianas de uso muy frecuente, así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato. Puede presentarse él mismo y ante otros, pedir y dar información personal básica sobre su domicilio, sus pertenencias y las personas que conoce. Puede relacionarse de forma elemental siempre que su interlocutor hable despacio y con claridad y esté dispuesto a cooperar.

4.2 Algunas características de los estudiantes que presentan la prueba

Se dará a conocer las características de los Factores Académicos, Demográficos y Socioeconómicos, pertenecientes a los estudiantes de Licenciatura en Matemáticas y Estadística. Los resultados se presentan en tablas las cuales incluyen, \bar{x} , valor promedio, DE , desviación estándar, CV , coeficiente de variación y porcentajes en cada categoría de respuesta de las variables.

Tabla 4-11: Características Factor Académico estudiantes Saber Pro.

Variable	Unidades de Medida/Categorías	Resumen Numérico
Tipo Bachille.	A:Académico	A:38 %
	T: Técnico	N:2 %
	N: Normalista Superior	T:60 %
Educ. Madre	N:Ninguno	N:1 %
	P:Postgrado	P: 1 %
	PI: Primaria incompleta	PI: 25 %
	PC: Primaria completa	PC: 26 %
	SI: Secundaria(bachillerato) Incompleta	SI: 20 %
	SC: Secundaria (bachillerato) completa	SC: 20 %
	ETI: Educación técnica o tecnológica incompleta	ETI: 2 %
	ETC: Educación técnica o tecnológica completa	ETC: 5 %
Educ. padre	N:Ninguno	N: 2 %
	P:Postgrado	P:1 %
	PI: Primaria incompleta	PI:27 %
	PC: Primaria completa	PC:27 %
	SI: Secundaria(bachillerato) incompleta	SI:22 %
	SC: Secundaria (bachillerato) completa	SC:14 %
	ETI: Educación técnica o tecnológica incompleta	ETI: 1 %
	ETC: Educación técnica o tecnológica completa	ETC: 3 %
	EPI: Educación profesional incompleta	EPI:1 %
	EPC: Educación profesional completa	EPC: 2 %

Tabla 4-12: Características Factor Demográfico estudiantes Saber Pro.

Variable	Unidad de medida/Categ.	Resumen Numérico
Estado civil	UL: Unión libre	UL: 9 %
	C: Casado	C: 18 %
	S: Soltero	S: 73 %
Género	F: Mujer	F: 42 %
	M: Hombre	M: 58 %
Situación del Hogar	H: Habitual o permanente	H: 76 %
	T: Temporal razones estudio	T: 24 %
Núm. de Personas Hogar		$\bar{x} = 4$
		$CV = 0,39$
		$DE = 1,650$
Cabeza de Familia	Si	Si: 79 %
	No	No: 21 %
Edad	Años Cumplidos	$\bar{x}=27$
		$CV=0.17$
		$DE=4.7$
Núm. de Personas a Cargo		$\bar{x} = 1$
		$CV : 1,46$
		$DE = 1,09$

Tabla 4-13: Características Factor Socioeconómico estudiantes Saber Pro.

Variable	Unidades de medida/Categorías	Resumen Numérico
Ocupación Madre	A:Administrador	A: 24 %
	O:Obrero	O: 6 %
	OR: Ocupación No Remunerada	OR: 70 %
Ocupación Padre	A:Administrador	A: 53 %
	O:Obrero	O: 22 %
	OR: Ocupación No Remunerada	OR: 11 %
	NT: No Trabajador Remunerado	NT: 14 %
Telefonía	Si	Si: 76 %
	No	No: 24 %
Número de dormitorios		$\bar{x} = 3$
		$CV = 0.40$
		$DE = 1.13$
No curso	NR: No Repaso	NR: 22 %
	RP: Repaso Cuenta Propia	RP: 55 %
	NA: Datos Faltantes	NA: 23 %
Valor matrícula	1:Menos de 500 mil	1: 69 %
	2: Entre 500 mil y menos de 1 millón	2: 26 %
	3:Entre 1 millón y 3 millones	3: 1 %
	4:Entre 3 millones y 5 millones	4: 1 %
	5:Más de 5 millones	5: 0 %
	6:Entre 500 mil y menos de 1 millón	6: 0 %
	7:No pago matrícula	7: 2 %
Internet	Si	Si: 35 %
	No	No: 65 %
Horas Trabajo		$\bar{x} = 16.4$
		$CV = 0,33$
		$DE = 5,48$

Tabla 4-14: Características Factor Socioeconómico estudiantes Saber Pro.

Variable	Unidades de medida/Categorías	Resumen Numérico
Estrato (Recibo Energía)	1: Estrato 1	1: 13 %
	2: Estrato 2	2: 59 %
	3: Estrato 3	3: 27 %
	4: Estrato 4	4: 1 %
Nivel Sisben	N1:Nivel 1	N1: 25 %
	N2:Nivel 2	N2: 49 %
	N3:Nivel 3	N3: 3 %
	NC: No Clasificado	NC: 22 %
	ON: Otro nivel	ON: 1 %
Celular	Si	Si: 96 %
	No	No: 4 %
TV	Si	Si: 47 %
	No	No: 53 %
Lavadora	Si	Si: 61 %
	No	No: 39 %
Ingreso Familiar mensual	1: 1 smmlv	1: 29 %
	2: 2y 3 menos smmlv	2: 59 %
	3: 3 y 4 menos smmlv	3: 10 %
	4: 4 y 5 menos smmlv	4: 1 %
Trabaja	A: Ayudante sin remuneración	A: 63 %
	SE: si, para tener experiencia	SE: 10 %
	SP: si, por ser plan de practica	SP: 1 %
	No	No: 26 %

De acuerdo a las Tablas 4-11, 4-12, 4-13, y 4-14, los estudiantes que presentan la prueba se identifican por tener una edad promedio de 27 años, el 58 % son hombres, cuyos padres tienen mayoritariamente (26 % y 27 %) grado escolar de primaria completa, con una ocupación no remunerada en el caso de las madres (70 %) y los padres con ocupación de administrador (53 %), siendo la mayoría solteros (73 %), con un pago de matrícula menor a 500,000 pesos (69 %), pertenecientes a un tipo de bachillerato normalista (60 %), con situación del hogar permanente (76 %), donde el promedio de personas en el hogar es de cuatro, caracterizados por ser cabeza de familia (79 %), con una persona en promedio a cargo, ubicados en estrato dos en el recibo de energía (59 %) y sisben (49 %), los cuales cuentan con celular (96 %), teléfono fijo (76 %) y lavadora

(61 %), de manera contraria no cuentan con internet (65 %), ni televisión (53 %), el ingreso familiar esta entre dos y tres salarios mínimos vigentes (59 %), tienen en promedio tres dormitorios en el hogar, se desempeñan laboralmente como ayudantes sin remuneración (63 %), trabajando en un promedio de 16 horas por semana, caracterizados por repasar por cuenta propia para la prueba Saber Pro (55 %).

A continuación se dará a conocer el análisis bidimensional o bivariado, donde se establecerán las relaciones entre dos variables, en específico entre los puntajes de las competencias genéricas (Inglés, Razonamiento Cuantitativo, Comunicación Escrita, Lectura crítica, Desempeño en Inglés y Desempeño en Comunicación Escrita) y los factores (Académico, Demográfico y Socioeconómico), buscando establecer las asociaciones pertinentes entre las variables de tipo cualitativas y cuantitativas. Se comenzará con las variables cualitativas usando el test Chi-cuadrado para demostrar estadísticamente las vinculaciones entre las variables, seguido del ANOVA de un factor, o test de Krukall-Wallis en el caso de las variables cuantitativas no cumplan normalidad, para comprobar diferencias con variables cuantitativas. Los soportes estadísticos donde se evidencian los siguientes resultados están ubicados en los Anexos ².

Al relacionar los puntajes obtenidos en los módulos del componente genérico y algunas características de los estudiantes se encontró asociación entre el desempeño de Comunicación Escrita y la educación de la madre ($p = 0.0192$), la educación del padre ($p = 0.0448$), el nivel de sisben ($p = 0.0013$), el porcentaje créditos cursados cuando presentó la prueba ($p = 3.887\text{e-}6$) y el haber tomado un refuerzo en competencias genéricas ($p = 9,15\text{e-}12$).

Se encontró una diferencia de medias de acuerdo al porcentaje de los créditos cursados en Comunicación Escrita ($p=0.038$), ver Tabla 4-15.

Tabla 4-15: Promedio en Comunicación EscritaI vs Porcentaje Créditos.

% Créditos	No sabe	Menos 75 %	75 %-80 %	81 %-90 %	Más 90 %
\bar{x}	11.20	11.25	10.30	10.10	9.80

También se encontró asociación entre el desempeño Inglés y la educación de la madre

² Anexo A

($p = 0.078$) y el haber tomado refuerzo en competencias genéricas ($p = 0.013$).

Se percibió una diferencia entre los promedios para el puntaje en Inglés según la educación de la madre ($p = 0.0064$), el pago de matrícula por cuenta de los padres y propio ($p = 0.02$ y $p = 0.009$) y el nivel del sisben ($p = 0.02$). También se observó una baja correlación inversa con la edad ($p = 0.036$), ver Tablas **4-16**, **4-17**, **4-18** y **4-19**, .

Tabla 4-16: Promedio en Inglés vs Educación madre.

Educación madre	ninguno	PI	PC	SI	SC	ETI	ETC	P
\bar{x}	9.0	9.50	9.60	9.90	9.80	9.85	11.0	8.9

Tabla 4-17: Promedio en Inglés vs Pago matrícula por padres.

Pago matrícula por padres	No	Si
\bar{x}	9.5	9.9

Tabla 4-18: Promedio en Inglés vs Pago matrícula propio.

Pago matrícula propio	No	Si
\bar{x}	10.75	9.6

Tabla 4-19: Promedio en Inglés vs Nivel sisben.

Nivel sisben	N1	N2	N3	Otro	No clasificado
\bar{x}	10.04	9.60	10.60	9.8	9.8

Respecto a Razonamiento Cuantitativo se encontró diferencia entre los promedios según sexo ($p = 0.02$), el uso de internet ($p = 0.002$) y el pago de la de matrícula por cuenta propia ($p = 0.01$), ver Tablas **4-20**, **4-21** y **4-22**.

Tabla 4-20: Promedio en Razonamiento Cuantitativo vs Pago matrícula propio.

Pago matrícula propio	No	Si
\bar{x}	11.65	10.40

Tabla 4-21: Promedio en Razonamiento Cuantitativo vs Internet.

Internet	No	Si
\bar{x}	10.2	10.8

Tabla 4-22: Promedio en Razonamiento Cuantitativo vs Sexo.

Sexo	M	H
\bar{x}	10.25	10.70

Se observó diferencia entre los promedios de Lectura Crítica según el nivel educativo de la madre ($p= 0.0008$), la ocupación del padre ($p= 9.62\text{e-}6$), y el pago de la matrícula por cuenta propia ($p= 0.0002$.)

5 FACTORES ASOCIABLES A LOS RESULTADOS DEL COMPONENTE GENÉRICO

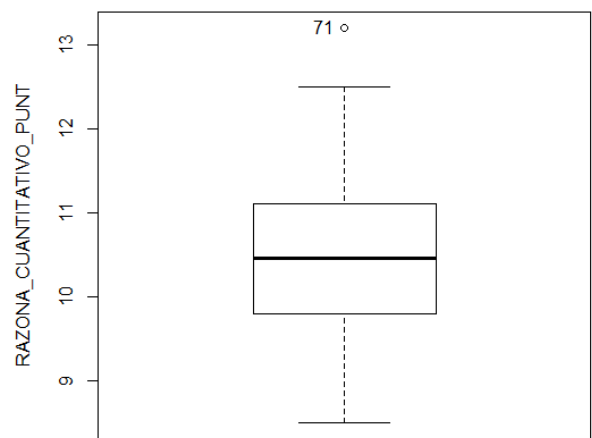
La construcción de los modelos se llevó a cabo a través de la especificación del modelo, selección (modelo parsimonioso), evaluación e interpretación.

5.1 ESPECIFICACIÓN DE LOS MODELOS

Se dará a conocer las distribuciones seleccionadas para los puntajes de Razonamiento Cuantitativo, Inglés, Lectura Critica y Comunicación Escrita, por medio de pruebas gráficas y del AIC (Criterio de Información Akaike), como también la exploración, a través de diagramas de caja, de presencia de datos atípicos.

5.1.1. Puntaje Razonamiento Cuantitativo

Figura 5-1: Datos atípicos en puntaje Razonamiento Cuantitativo.



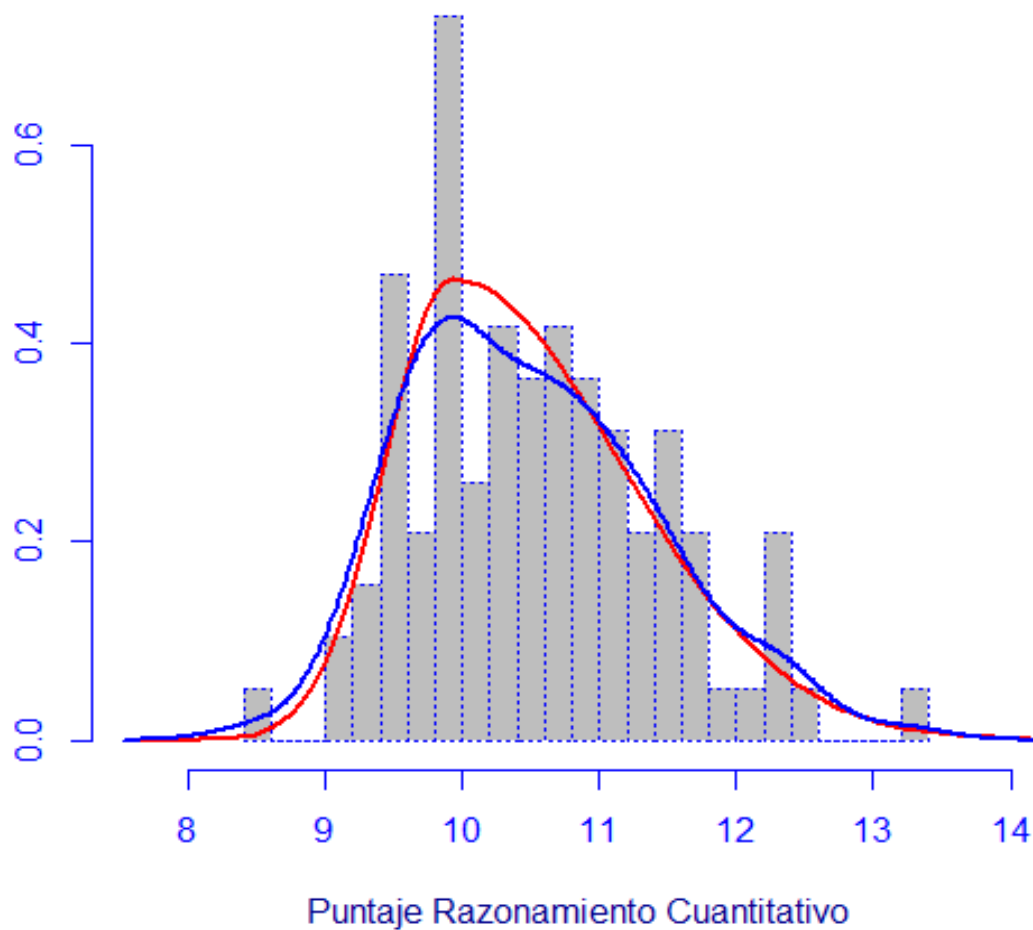
En la figura 5-1 se observa un dato atípico obtenido por el estudiante número 71, con un valor de 13.2 en el puntaje de Razonamiento Cuantitativo, clasificado como un estudiante que obtuvo un puntaje excelente.

Figura 5-2: Selección de distribución para puntaje Razonamiento Cuantitativo.

SN2	SN1	IGAMMA	RG	ST3	SEP3	IG	GG
249.1446	250.1482	250.6090	251.0922	251.1440	251.1442	251.3538	251.3669
LOGNO	SHASHo	BCCGo	SEP1	ST2	ST1	GA	BCPEo
251.3721	251.4681	251.4698	252.1433	252.1482	252.1482	252.3214	252.3352
SEP4	exGAUS	JSU	ST5	EGB2	BCTo	PE	TF
252.6517	252.9619	253.2359	253.3327	253.3351	253.4698	256.5004	256.8121
SEP2	LO	ST4	GT	WEI3	GU	EXP	PARETO2
256.8633	257.1448	257.1584	257.3337	273.0743	282.7374	645.5402	647.5404

De acuerdo a los valores del AIC, la distribución que describe el comportamiento del puntaje de Razonamiento Cuantitativo es Skew Normal 2 (SN2).

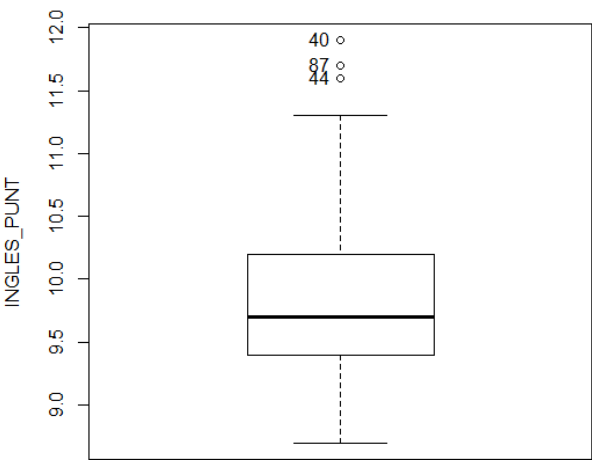
Figura 5-3: Ajuste puntaje Razonamiento Cuantitativo.



En la figura **5-3** se muestra el ajuste del puntaje a la distribución seleccionada, donde la línea roja corresponde a la distribución SN2 y la línea azul al puntaje de Razonamiento Cuantitativo.

5.1.2. Puntaje Inglés

Figura 5-4: Datos atípicos en puntaje Inglés.

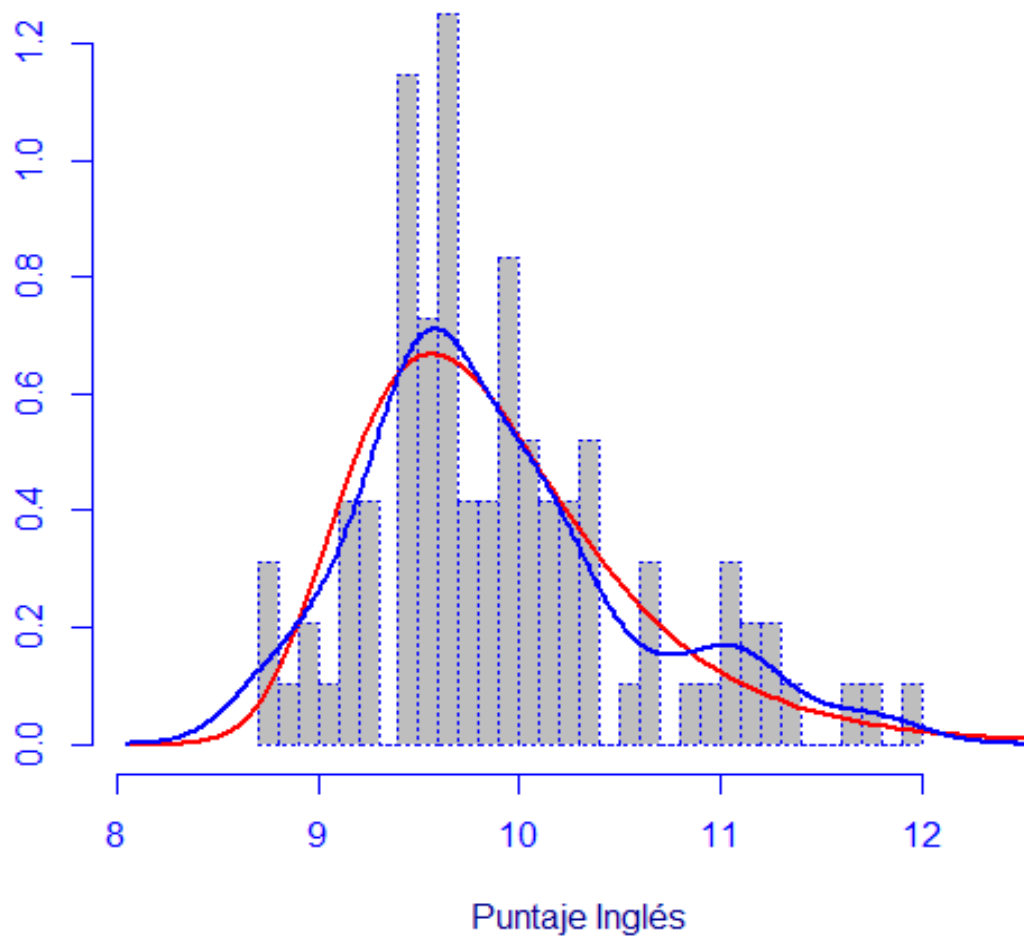


En la figura 5-4 se presentan datos atípicos por los estudiantes número 40, 44 y 87 con un valores de 11.7, 11.9 y 14 en el puntaje de Inglés, clasificados como estudiantes que obtuvieron puntajes excelentes.

Figura 5-5: Selección de distribución para puntaje Inglés.

RG	SEP2	exGAUS	SN1	SEP3	GG	SHASHo	BCCGo	SN2
190.6499	191.7546	192.1966	192.4859	192.6415	192.6854	192.7968	193.0557	193.7232
EGB2	SEP1	SEP4	ST1	ST2	JSU	ST5	BCPEo	BCTo
193.8315	193.8649	193.8971	194.4691	194.4792	194.6385	194.7044	195.0473	195.0557
ST3	IGAMMA	LOGNO	IG	ST4	GA	LO	PE	TF
195.4029	197.1747	198.5071	198.5087	198.8732	199.9748	201.5827	203.4462	203.9640
GT	WEI3	GU	EXP	PARETO2				
205.4462	229.1260	238.5088	633.7825	635.7827				

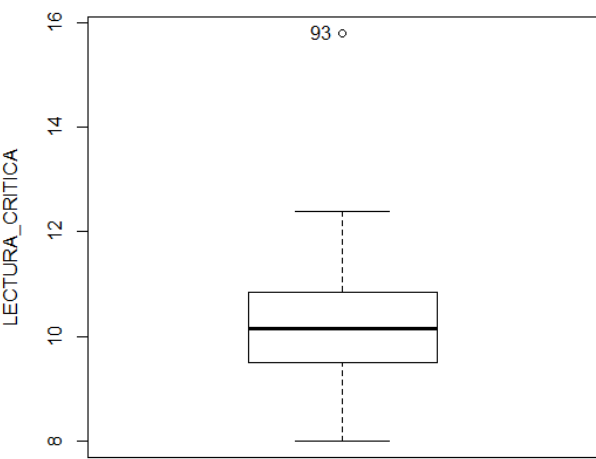
De acuerdo alos valores del AIC, la distribución que describe el comportamiento del puntaje de Inglés es Reverse Gumbel (RG).

Figura 5-6: Ajuste puntaje de Inglés.

En la figura **5-6** se muestra el ajuste del puntaje a la distribución seleccionada, donde la línea roja corresponde a la distribución RG y la línea azul al puntaje de Inglés.

5.1.3. Puntaje Lectura Crítica

Figura 5-7: Datos atípicos en puntaje Lectura Crítica.

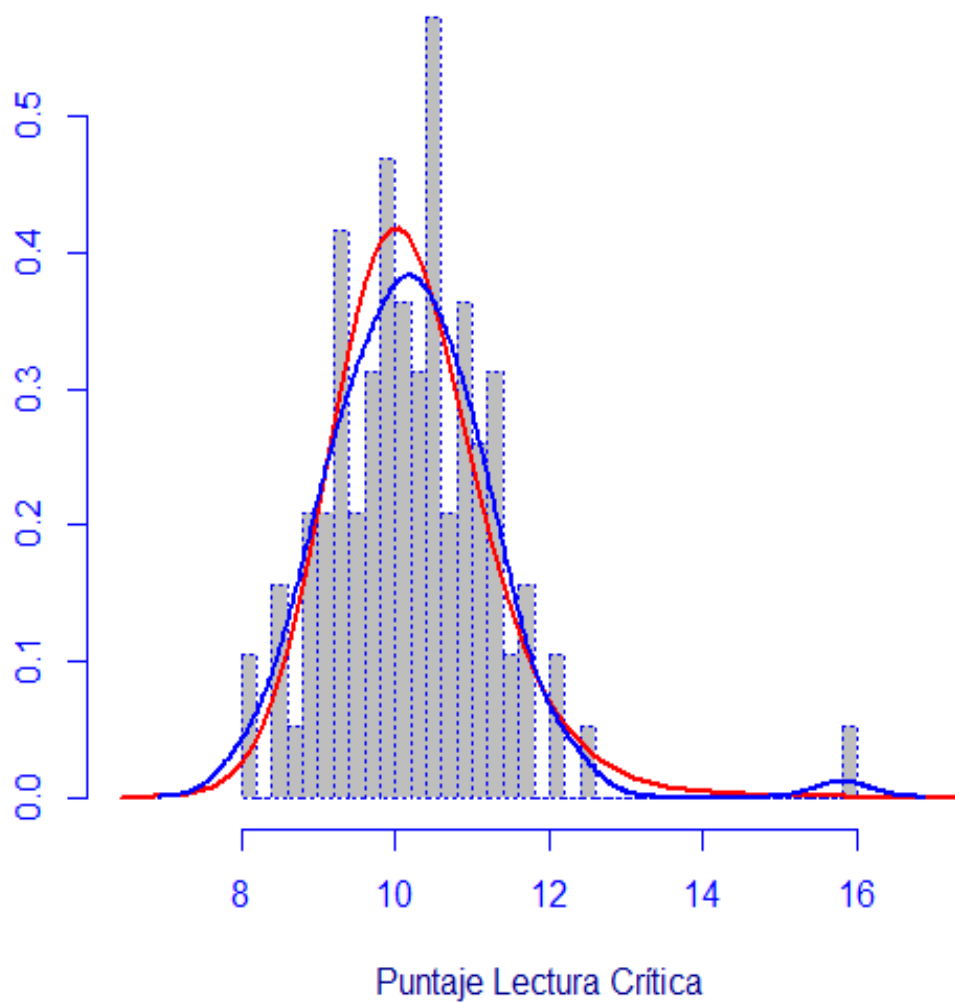


En la figura 5-7 se presenta un dato atípico por el estudiante número 93 con un valor de 15.8 en el puntaje de Lectura Crítica, clasificado como un estudiante que obtuvo un puntaje excelente.

Figura 5-8: Selección de distribución para puntaje Lectura Crítica.

exGAUS	ST4	GT	LO	TF	IGAMMA	BCTo	ST1	BCCGo
280.9632	281.2573	281.6129	281.7255	282.0217	282.1327	282.2924	282.4549	282.5836
ST5	GG	ST2	BCPEo	JSU	ST3	LOGNO	EGB2	IG
282.6258	282.8522	283.0677	283.3182	283.3643	283.5094	283.6734	283.7478	283.8704
GIG	RG	SEP4	GA	SEP1	PE	SN1	SHASHo	SEP3
284.1327	284.5162	284.9279	285.9256	286.2517	286.2524	286.2686	286.5347	286.6644
SEP2	SN2	WEI3	GU	EXP	PARETO2			
287.5574	288.3647	329.9351	360.3500	639.8592	641.8593			

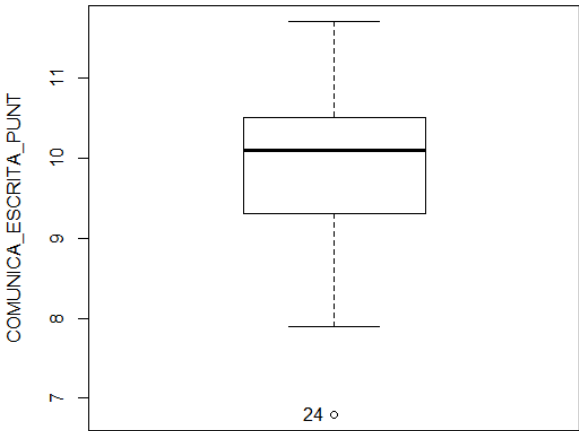
De acuerdo a los valores del AIC, la distribución que describe el comportamiento del puntaje de Lectura Crítica es Exponencial Gaussian (exGaus).

Figura 5-9: Ajuste puntaje Lectura Crítica.

En la figura 5-9 se muestra el ajuste del puntaje a la distribución seleccionada, donde la línea roja corresponde a la distribución SN2 y la línea azul al puntaje de Razonamiento Cuantitativo.

5.1.4. Puntaje Comunicación Escrita

Figura 5-10: Datos atípicos en puntaje Comunicación Escrita.

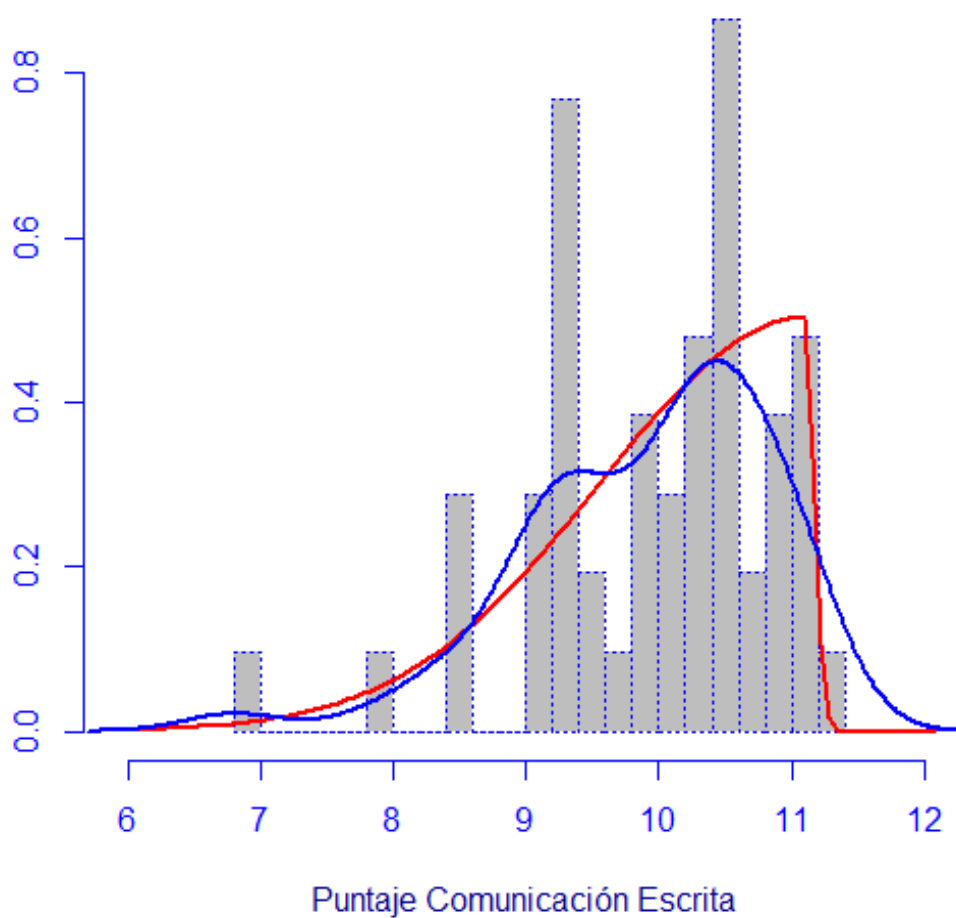


En la figura 5-10 se presenta un dato atípico por el estudiante número 24 con un valor de 15.8 en el puntaje de Comunicación Escrita, clasificado como excelente.

Figura 5-11: Selección de distribución para puntaje Comunicación Escrita.

SN2	SEP2	SEP1	SEP3	ST3	ST2	ST1	GU
129.5119	131.3048	131.4166	131.4226	131.5119	131.6253	131.6253	132.0039
GG	BCPEo	WEI3	BCCGo	SHASHo	JSU	BCTo	EGB2
132.5433	132.7331	132.9574	133.2149	133.5567	134.8901	135.2149	135.3879
ST5	GT	LO	ST4	TF	PE	SN1	exGAUS
135.6832	138.9244	140.7279	141.0909	142.2961	143.3810	143.4774	143.4814
GA	LOGNO	IG	IGAMMA	RG	EXP	PARETO2	
145.1051	147.2078	147.3215	149.5840	164.6751	344.6456	346.6457	

De acuerdo a los resultados del AIC, la distribución que describe el comportamiento del puntaje de Comunicación Escrita es Skew Normal 2 (SN2).

Figura 5-12: Ajuste puntaje Comunicación Escrita.

5.2 COMPONENTE SISTEMÁTICO

A continuación se pondrá en conocimiento las variables explicativas fijas, medidas sin error respecto a los factores Académicos, Demográficos y Socioeconómicos, cuya notación corresponde a:

Tabla 5-1: Notación de variables explicativas del factor Académico.

VARIABLES	DESCRIPCIÓN
x_1	Tipo de bachillerato
x_2	Nivel educativo de la madre
x_3	Nivel educativo del padre
x_4	No tomar curso

Tabla 5-2: Notación de variables explicativas de los factor Demográfico.

VARIABLES	DESCRIPCIÓN
x_1	Número de personas a cargo
x_2	Número de personas en el hogar
x_3	Edad
x_4	Sexo
x_5	Cabeza de familia
x_6	Estado civil
x_7	Situación del hogar

Tabla 5-3: Notación de variables explicativas de los factor Socioeconómico.

VARIABLES	DESCRIPCIÓN
x_1	Número de dormitorios de la residencia
x_2	Horas de trabajo
x_3	Ocupación de la madre
x_4	Ocupación del padre
x_5	Estrato de residencia factura energía
x_6	Recursos forma de pago
x_7	Celular
x_8	Internet
x_9	Televisor
x_{10}	Telefono fijo
x_{11}	Computador
x_{12}	Lavadora
x_{13}	Salario mensual familiar
x_{14}	Trabaja actualmente
x_{15}	Valor anual matricula

Los Modelos Aditivos Generalizados de Ubicación, Escala y Forma para los puntajes de Razonamiento Cuantitativo, Inglés, Lectura Crítica y Comunicación Escrita quedan determinados por los siguientes predictores lineales:

Factor Académico:

$$\eta_i = \beta_0 + \sum_{j=1}^4 \tau_{jm} X_{ij} \quad (5-1)$$

Factor Demográfico:

$$\eta_i = \beta_0 + \sum_{j=1}^3 \beta_j X_{ij} + \sum_{j=4}^7 \psi_{jn} X_{ij} \quad (5-2)$$

Factor Socioeconómico:

$$\eta_i = \beta_0 + \sum_{j=1}^2 \beta_j X_{ij} + \sum_{j=3}^{15} \delta_{jp} X_{ij} \quad (5-3)$$

El predictor lineal expresa los efectos de las variables explicativas de la siguiente manera:

- τ y β (escalares) son el efecto de las variables del factor Académico.

β_0 = Intercepto

τ_{1m} = Efecto del tipo de bachillerato, $m = 1, 2, 3, 4$.

τ_{2m} = Efecto del nivel educativo de la madre, con $m = 1, 2, 3, 4, 5$.

τ_{3m} = Efecto del nivel educativo del padre, con $m = 1, 2, 3, 4, 5$.

τ_{4m} = Efecto de no tomar curso, con $m = 1, 2$.

- ψ y β (escalares) son el efecto de las variables del factor Demográfico.

β_0 = Intercepto

β_j = Efecto de X_{ij} , con $j = 1, 2, 3$.

ψ_{4n} = Efecto del sexo, con $n = 1, 2$.

ψ_{5n} = Efecto de ser cabeza de familia, con $n = 1, 2$.

ψ_{6n} = Efecto del estado civil, con $n = 1, 2, 3, 4, 5$.

ψ_{7n} = Efecto de la situación del hogar, con $n = 1, 2$.

- δ y β (escalares) son el efecto de las variables del factor Socioeconómico.

β_0 = Intercepto

β_j = Efecto de X_{ij} , con $i = 1, 2$.

δ_{3p} = Efecto de la ocupación de la madre, con $p = 1, 2, 3, 4, 5$.

δ_{4p} = Efecto de la ocupación del padre, con $p = 1, 2, 3, 4, 5$.

δ_{5p} = Efecto del estrato de residencia factura de energia, con $p = 1, 2, 3, 4, 5, 6, 7$.

δ_{6p} = Efecto de los recursos de pago, con $p = 1, 2$.

δ_{7p} = Efecto del celular, con $p = 1, 2$.

δ_{8p} = Efecto del internet, con $p = 1, 2$.

δ_{9p} = Efecto del televisor, con $p = 1, 2$.

δ_{10p} = Efecto del telefono fijo, con $p = 1, 2$.

δ_{11p} = Efecto del computador, con $p = 1, 2$.

δ_{12p} = Efecto de la lavadora, con $p = 1, 2$.

δ_{13p} = Efecto del salario mensual familiar, con $p = 1, 2, 3, 4, 5, 6, 7$.

δ_{14p} = Efecto del trabajo, con $p = 1, 2, 3, 4$.

δ_{15p} = Efecto del valor anual de la matricula, con $p = 1, 2, 3, 4, 5, 6$.

5.3 SELECCIÓN, EVALUACIÓN E INTERPRETACIÓN DEL MODELO ÓPTIMO

Se dará a conocer los modelos óptimos para los puntajes de Razonamiento Cuantitativo, Inglés, Lectura Crítica y Comunicación Escrita mediante los factores Académicos, Demográficos y Socioeconómicos, donde su selección está dada por el AIC (Criterio de Información Akaike). Además se proporcionará los diagnósticos sobre el ajuste de cada distribución seleccionada para los puntajes. Seguido de las interpretaciones del promedio, μ , la varianza, σ y la asimetría, ν . Valores y detección de atípicos en las variables explicativas cuantitativas (edad, número de personas a cargo y número de personas en el hogar). También se dará a conocer los modelos del desempeño en Comunicación Escrita e Inglés de acuerdo a los factores nombrados anteriormente.

5.3.1. Puntaje Razonamiento Cuantitativo

A continuación se presentarán los modelos correspondientes a los factores Académicos, Demográficos y Socioeconómicos del puntaje de Razonamiento Cuantitativo.

Modelo Factor Académico

La Figura 5-13 presenta el modelo óptimo correspondiente al Factor Académico.

Figura 5-13: Modelo óptimo Razonamiento Cuantitativo Factor Académico.

```
*****
Family:  c("SN2", "skew normal type 2")

Call:  gamlss(formula = RAZO_CUANTITATIVO_PUNT ~ EDUCAMADRE1 +
  EDUCAPADRE1 + NOCURSO, sigma.formula = ~NOCURSO,
  nu.formula = ~EDUCAMADRE1 + EDUCAPADRE1 + NOCURSO,
  family = SN2, data = na.omit(Dataset), trace = FALSE)

Fitting method: RS()

-----
Mu link function:  identity
Mu Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    14.32849    1.26502   11.327 4.72e-14 ***
EDUCAMADRE1Primaria  0.02456    0.76069    0.032  0.974
EDUCAPADRE1PRIMARIA -4.55260    0.47664   -9.551 7.13e-12 ***
EDUCAPADRE1SECUNDARIA -4.78265    0.71478   -6.691 5.06e-08 ***
NOCURSORepaso por cuenta propia  1.10659    0.69815    1.585  0.121
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function:  log
Sigma Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.4003    0.7915  -0.506  0.616
NOCURSORepaso por cuenta propia  0.5320    0.8072    0.659  0.514
---

-----
Nu link function:  log
Nu Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)     5.31761    1.41617    3.755 0.000552 ***
EDUCAMADRE1Primaria  0.06008    0.56827    0.106 0.916335
EDUCAPADRE1PRIMARIA -4.72253    0.33414  -14.134 < 2e-16 ***
EDUCAPADRE1SECUNDARIA -4.46492    0.59471   -7.508 3.72e-09 ***
NOCURSORepaso por cuenta propia -0.76878    0.95478   -0.805 0.425469
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  52
Degrees of Freedom for the fit:  12
Residual Deg. of Freedom:  40
                        at cycle: 20

Global Deviance:    141.7249
AIC:                165.7249
SBC:                189.1398
*****
```

Figura 5-14: Diagnóstico de la distribución modelo Razonamiento Cuantitativo Factor Académico.

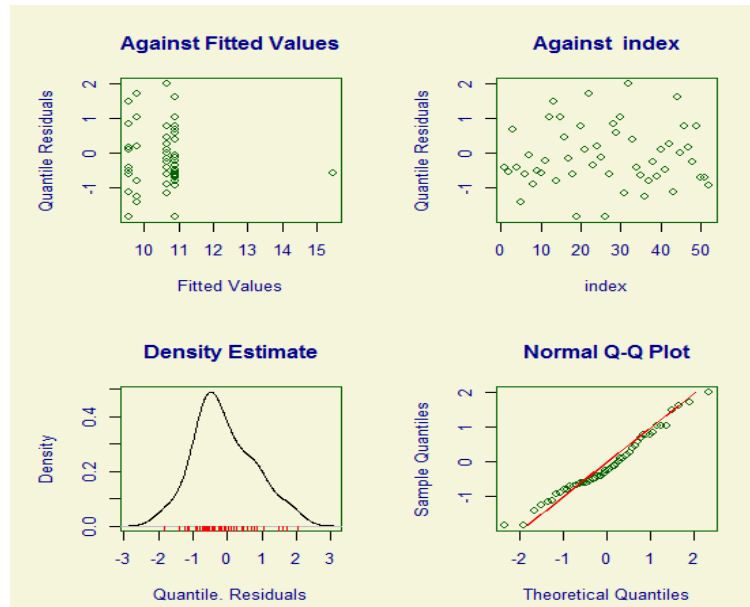


Figura 5-15: Prueba ajuste modelo Razonamiento Cuantitativo Factor Académico.

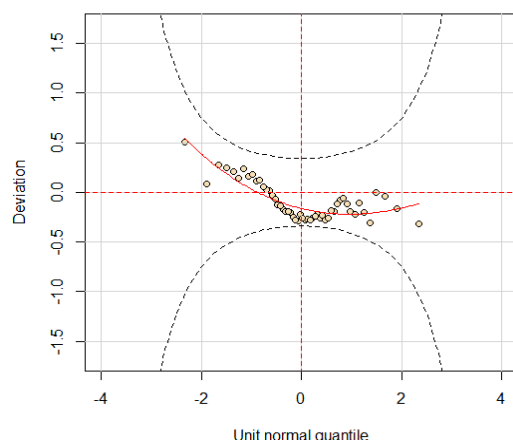
```

*****
Summary of the Quantile Residuals
      mean   = -0.09469325
    variance = 0.7674289
  coef. of skewness = 0.3858878
    coef. of kurtosis = 2.66799
Filliben correlation coefficient = 0.9879673
*****

```

En las figuras 5-14 y 5-15 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-16: Ajuste de la distribución modelo Razonamiento Cuantitativo Factor Académico.



A partir de la validación del modelo óptimo se puede decir que, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo la educación del padre se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes donde el padre tiene educación primaria, el puntaje promedio de Razonamiento Cuantitativo disminuye 4.72 puntos respecto a los estudiantes con padre sin ningún tipo de estudios. Con un nivel de significancia inferior al 5 %, que aquellos estudiantes donde el padre tiene educación secundaria, el puntaje promedio de Razonamiento Cuantitativo disminuye 4.46 puntos respecto a los estudiantes con padre sin ningún tipo de estudios.

También se determina que, la variable del factor académico (Tomar curso-repaso por cuenta propia) no explica la variabilidad del puntaje de Razonamiento Cuantitativo de estudiantes Licenciatura en Matemáticas y Estadística. Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo la educación del padre se tiene que, con un nivel de significancia inferior al 5 %, que la explicación del sesgo en el puntaje de Razonamiento Cuantitativo se da por la educación primaria y secundaria del padre respecto a los estudiantes con padre sin ningún tipo de estudios.

Modelo Factor Demográfico

La Figura 5-17 presenta el modelo óptimo correspondiente al Factor Demográfico.

Figura 5-17: Modelo óptimo Razonamiento Cuantitativo Factor Demográfico.

```
*****
Family: c("SN2", "skew normal type 2")

Call: gamlss(formula = RAZO_CUANTITATIVO_PUNT ~ GÉNERO + CABEZAFMLIA + EDAD + SITUACIÓNHOGAR + NÚMERO.PERSOS.HOGAR + NUMERO.DE.PERSOS.A.CARGO,
  sigma.formula = ~GÉNERO + CABEZAFMLIA + EDAD + SITUACIÓNHOGAR + NÚMERO.PERSOS.HOGAR + NUMERO.DE.PERSOS.A.CARGO, nu.formula = ~ESTADO.CIVIL +
  GÉNERO + CABEZAFMLIA + EDAD + SITUACIÓNHOGAR + NÚMERO.PERSOS.HOGAR + NUMERO.DE.PERSOS.A.CARGO, family = SN2, data = na.omit(Dataset),
  trace = FALSE)

Fitting method: RS()

-----
Mu link function: identity
Mu Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.97033    0.24003   49.871 < 2e-16 ***
GÉNEROM         -0.98771    0.09015  -10.956 2.75e-14 ***
CABEZAFMLIAasi  -1.14683    0.18488   -6.203 1.56e-07 ***
EDAD           -0.02372    0.01018   -2.330  0.0244 *
SITUACIÓNHOGAR(temporal por razones de estudio) -0.32859    0.12981   -2.531  0.0149 *
NÚMERO.PERSOS.HOGAR -0.11948    0.05139   -2.325  0.0247 *
NUMERO.DE.PERSOS.A.CARGO  0.43603    0.07990   5.458 1.98e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function: log
Sigma Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)     4.42088    1.15843   3.816 0.000411 ***
GÉNEROM          0.36693    0.32041   1.145 0.258185
CABEZAFMLIAasi  -1.04026    0.48125  -2.162 0.036007 *
EDAD            -0.14329    0.03987  -3.594 0.000804 ***
SITUACIÓNHOGAR(temporal por razones de estudio) -0.26212    0.34212  -0.766 0.447575
NÚMERO.PERSOS.HOGAR -0.47102    0.10739  -4.386 6.88e-05 ***
NUMERO.DE.PERSOS.A.CARGO  0.56035    0.18714   2.994 0.004458 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function: log
Nu Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.284149    1.616945  -0.176  0.8613
ESTADO.CIVILcasado -0.783148    0.605296  -1.294  0.2026
ESTADO.CIVILSoltero -1.082154    0.614379  -1.761  0.0853 .
GÉNEROM         1.744792    0.278735   6.260 1.53e-07 ***
CABEZAFMLIAasi  0.330750    0.622296   0.532  0.5978
EDAD            0.004728    0.050806   0.093  0.9263
SITUACIÓNHOGAR(temporal por razones de estudio) -0.182377    0.339230  -0.538  0.5936
NÚMERO.PERSOS.HOGAR  0.086629    0.122433   0.708  0.4830
NUMERO.DE.PERSOS.A.CARGO -0.115188    0.222474  -0.518  0.6073
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 52
Degrees of Freedom for the fit: 23
Residual Deg. of Freedom: 29
      at cycle: 20

Global Deviance: 79.63346
AIC: 125.6335
SBC: 170.5121
```

Figura 5-18: Diagnóstico de la distribución modelo Razonamiento Cuantitativo Factor Demográfico.

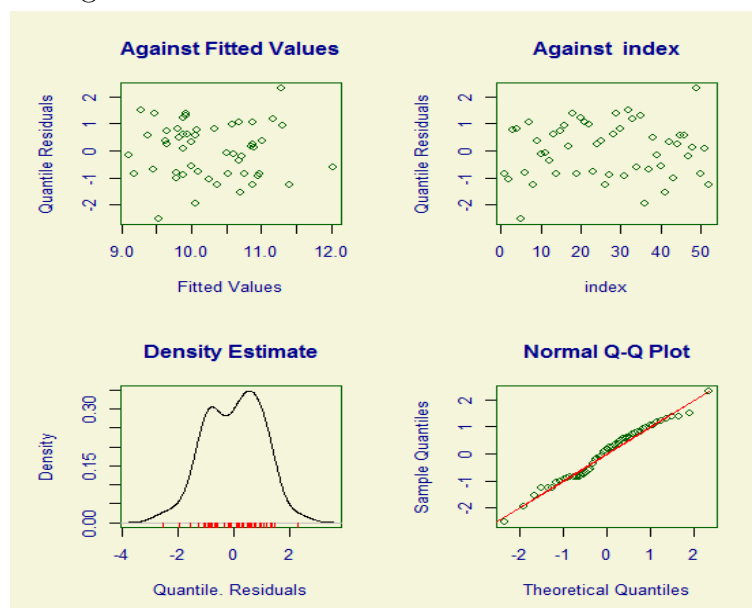


Figura 5-19: Prueba ajuste modelo Razonamiento Cuantitativo Factor Demográfico.

```
*****
Summary of the Quantile Residuals
      mean   = 0.01733885
      variance = 1.0062
      coef. of skewness = -0.1767553
      coef. of kurtosis  = 2.486836
Filliben correlation coefficient = 0.9898515
*****
```

En las figuras 5-18 y 5-19 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-20: Ajuste de la distribución modelo Razonamiento Cuantitativo Factor Demográfico.

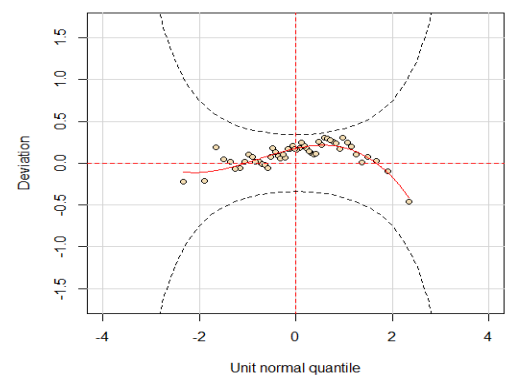


Figura 5-21: Valores Atípicos Número de personas a cargo Razonamiento Cuantitativo.

```
> Q.stats(w1,xvar=Dataset$NUMERO.DE.PERSOS.A.CARGO,n.inter=4)
```

	Z1	Z2	Z3	Z4	AgostinoK2	N
-0.5 to 0.5	-0.53975549	-0.8686339	-0.02583435	-1.2486020	1.5596743	55
0.5 to 1.5	0.76462780	1.5990259	-1.19301353	0.9682379	2.3607659	22
1.5 to 5.5	-0.02687775	-0.3253942	1.30038325	-0.2728947	1.7654681	19
TOTAL Q stats	0.87671407	3.4172901	3.11494529	2.5709630	5.6859083	96
df for Q stats	1.00000000	2.00000000	2.00000000	3.00000000	5.00000000	0
p-val for Q stats	0.34910325	0.1811110	0.21066783	0.4626029	0.3379906	0

Figura 5-22: Valores Atípicos Edad Razonamiento Cuantitativo.

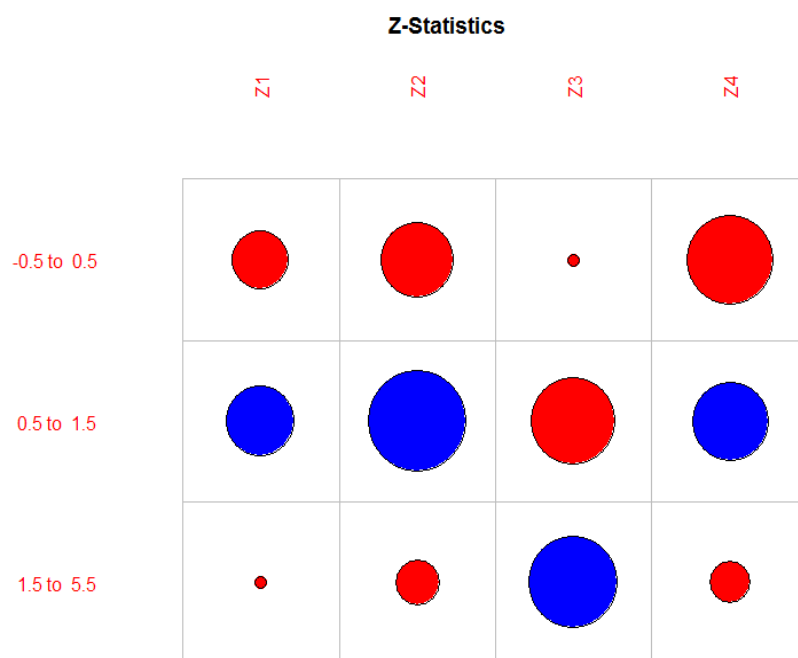
```
> Q.stats(w1,xvar=Dataset$EDAD,n.inter=4)
```

	Z1	Z2	Z3	Z4	AgostinoK2	N
19.5 to 24.5	0.3660966	-0.1127897	0.5497000	-1.4012342	2.2656274	33
24.5 to 26.5	-0.8556175	1.0313549	-1.2296773	0.2968796	1.6002438	22
26.5 to 29.5	0.5587904	-0.2965657	-0.2153982	-0.4944104	0.2908380	18
29.5 to 51.5	-0.2378101	-0.4360240	1.8967170	1.6676486	6.3785874	23
TOTAL Q stats	1.2349084	1.3544825	5.4582083	5.0770883	10.5352966	96
df for Q stats	2.00000000	3.00000000	3.00000000	4.00000000	7.00000000	0
p-val for Q stats	0.5393157	0.7162381	0.1411597	0.2794788	0.1602127	0

Figura 5-23: Valores atípicos Número personas hogar Razonamiento Cuantitativo.

```
> Q.stats(w2,xvar=Dataset$NÚMERO.PERSOS.HOGAR ,n.inter=4)
```

	Z1	Z2	Z3	Z4	AgostinoK2	N
0.5 to 3.5	0.3474194	1.4994118	-0.6373499	0.3902660	0.5585225	35
3.5 to 4.5	-0.7430257	-0.2523985	0.7498426	-2.0041121	4.5787295	23
4.5 to 6.5	0.6196783	-1.0390885	-0.1190100	0.5268807	0.2917667	30
6.5 to 9.5	-0.8639611	-1.1090261	-0.6045277	0.1275054	0.3817113	8
TOTAL Q stats	1.8032174	4.6215844	1.3480960	4.4626340	5.8107300	96
df for Q stats	2.0000000	3.0000000	3.0000000	4.0000000	7.0000000	0
p-val for Q stats	0.4059161	0.2016983	0.7177451	0.3470011	0.5620153	0

Figura 5-24: Prueba de valores atípicos número de personas a cargo Razonamiento Cuantitativo.

Las variables explicativas número de personas a cargo y edad no presentan datos atípicos.

Figura 5-25: Prueba de valores atípicos Edad Razonamiento Cuantitativo.

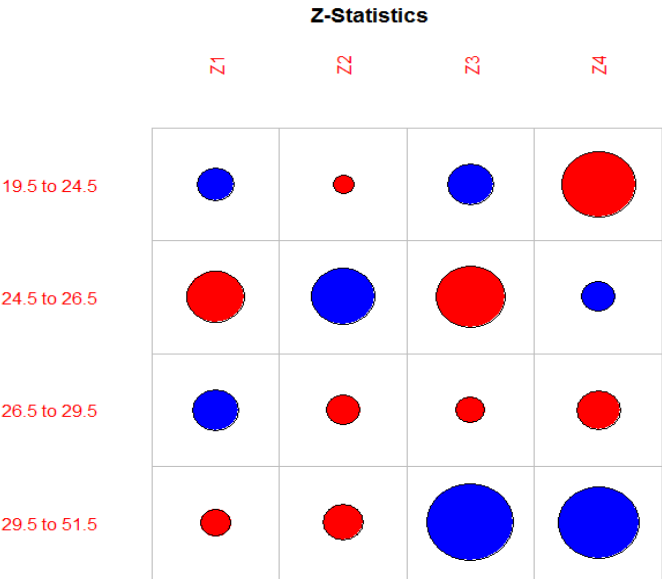
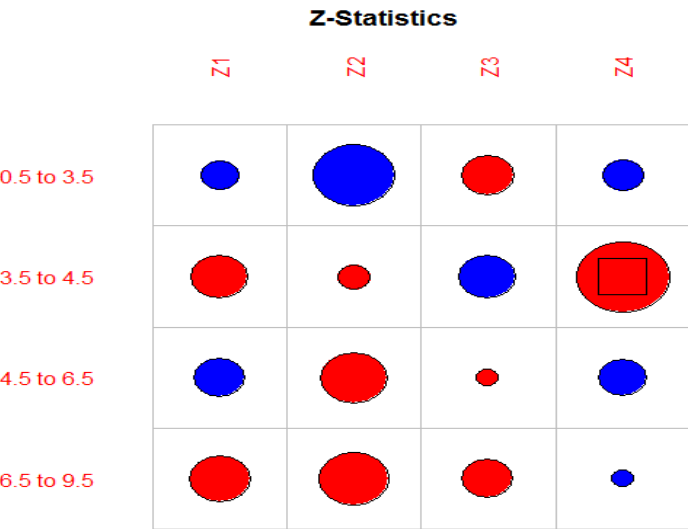


Figura 5-26: Prueba de valores atípicos Número de personas en el hogar Razonamiento Cuantitativo.



En la variable número de personas en el hogar se presenta un dato atípico con un valor de -2.0, indicando un bajo número de personas en el hogar del estudiante.

A partir de la validación del modelo óptimo se puede decir que, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el sexo, se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes con sexo masculino, el puntaje promedio de Razonamiento disminuye 0.98 puntos respecto al sexo femenino.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el hecho de ser cabeza de familia, se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes que son cabeza de familia, el puntaje promedio de Razonamiento Cuantitativo disminuye en 1.14 puntos respecto a los estudiantes que no lo son.

Al comparar dos estudiantes donde el segundo difiere del primero en tan solo un año de su edad, se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que presenta mayor edad, obtienen en promedio una disminución de 0.02 puntos en el puntaje de Razonamiento Cuantitativo.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo la situación del hogar, se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes donde la situación del hogar es temporal por razones de estudio, el puntaje promedio de Razonamiento Cuantitativo disminuye en 0.32 puntos respecto a los estudiantes con situación del hogar habitual o permanente.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo una persona de más en el hogar, se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que presentan un mayor número de personas en el hogar, obtienen en promedio una disminución de 0.12 puntos en el puntaje de Razonamiento Cuantitativo teniendo en cuenta alguna de las variables de estudio.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo una persona de más a cargo, se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que tienen personas a cargo, obtienen en promedio un aumento de 0.43 puntos en el puntaje de Razonamiento Cuantitativo.

También se determina que, al comparar dos estudiantes donde el segundo difiere del primero en tan solo un año de edad, se tiene que, con un nivel de significancia del 5 %, aquellos estudiantes de Licenciatura en Matemáticas y Estadística que presentan mayor edad, obtienen una mayor variabilidad de 2.82 puntos en el puntaje de Razonamiento Cuantitativo.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo una persona de más en el hogar, se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que presentan un mayor número de personas en el hogar, obtienen una mayor variabilidad de 0.47 puntos en el puntaje de Razonamiento Cuantitativo.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo una persona de más a cargo, se tiene que, con un nivel de significancia del 1 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que tienen personas a cargo, obtienen una menor variabilidad de 1.75 puntos en el puntaje de Razonamiento Cuantitativo.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo en el sexo se tiene que, con un nivel de significancia del 10 %, que la explicación del sesgo en el puntaje de Razonamiento Cuantitativo se da por el sexo masculino de los estudiantes respecto al sexo femenino.

Modelo Factor Socioeconómico

La Figura 5-27 presenta el modelo óptimo correspondiente al Factor Socioeconómico.

Figura 5-27: Modelo óptimo Razonamiento Cuantitativo Factor Socioeconómico.

```
*****
Family:  c("SN2", "skew normal type 2")

Call:  gamlss(formula = RAZO_CUANTITATIVO_PUNT ~ OCUP_MADRE1 + OCUP_PADRE1 + Recursos_forma.de.pago +
  CELULAR + INTERNET + TV + TELEFONIA + COMPUTADOR + LAVADORA + FMIAR_MENSUAL + NÚMERODORMITORIOS
  sigma.formula = ~1, nu.formula = ~1, family = SN2, data = na.omit(Dataset))

Fitting method: RS()

-----
Mu link function:  identity
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-193.300	294.207	-0.657	0.5169
OCUP_MADRE1Obrero	143.783	88.116	1.632	0.1148
OCUP_MADRE1Ocupación no remunerado	81.633	77.034	1.060	0.2990
OCUP_PADRE1No trabajador remunerado	188.928	211.604	0.893	0.3801
OCUP_PADRE1Obrero	-77.779	54.761	-1.420	0.1674
OCUP_PADRE1Ocupación no remunerado	-12.040	59.047	-0.204	0.8400
Recursos_forma.de.pagosi	14.002	49.360	0.284	0.7789
ESTRATORECIBOENERGIAESTR2	-13.902	94.764	-0.147	0.8845
ESTRATORECIBOENERGIAESTR3	22.977	85.968	0.267	0.7914
ESTRATORECIBOENERGIAESTR4	213.172	211.818	1.006	0.3235
VALR_MATRICULA_AÑO.ANTERIOR(Entre 3 millones y 5 millones\r\r\n)	417.944	261.906	1.596	0.1226
VALR_MATRICULA_AÑO.ANTERIOR(Entre 500 mil y menos de 1 millón)	122.620	234.077	0.524	0.6048
VALR_MATRICULA_AÑO.ANTERIOR(Más de 5 millones\r\r\n)	128.842	269.011	0.479	0.6360
VALR_MATRICULA_AÑO.ANTERIOR(Menos de 500 mil\r\r\n)	135.213	227.833	0.593	0.5580
CELULARsi	-20.280	83.870	-0.242	0.8108
INTERNETsi	11.295	63.295	0.178	0.8598
TVsi	37.059	72.350	0.512	0.6128
TELEFONIASi	-47.065	51.818	-0.908	0.3721
COMPUTADORsi	133.854	104.381	1.282	0.2110
LAVADORAsi	-125.003	86.467	-1.446	0.1602
FMIAR_MENSUAL{3 y 4 menos sm}	7.621	103.076	0.074	0.9416
FMIAR_MENSUALism	71.103	96.459	0.737	0.4676
NÚMERODORMITORIOS	8.223	29.379	0.280	0.7818
HORAS_TRABAJO	-2.849	4.597	-0.620	0.5409
TRABAJA{si, para tener experiencia}	-54.426	77.818	-0.699	0.4905
TRABAJA{si, por ser plan de practica}	-182.741	103.701	-1.762	0.0898

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function:  log
Sigma Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.30322	0.04031	106.8	<2e-16 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function:  log
Nu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.24004	0.03785	-6.342	5.96e-08 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  52
Degrees of Freedom for the fit:  28
Residual Deg. of Freedom:  24
at cycle:  20

Global Deviance:  599.0686
AIC:  655.0686
SBC:  709.7034
*****
```


Figura 5-28: Diagnóstico de la distribución modelo Razonamiento Cuantitativo Factor Socioeconómico.

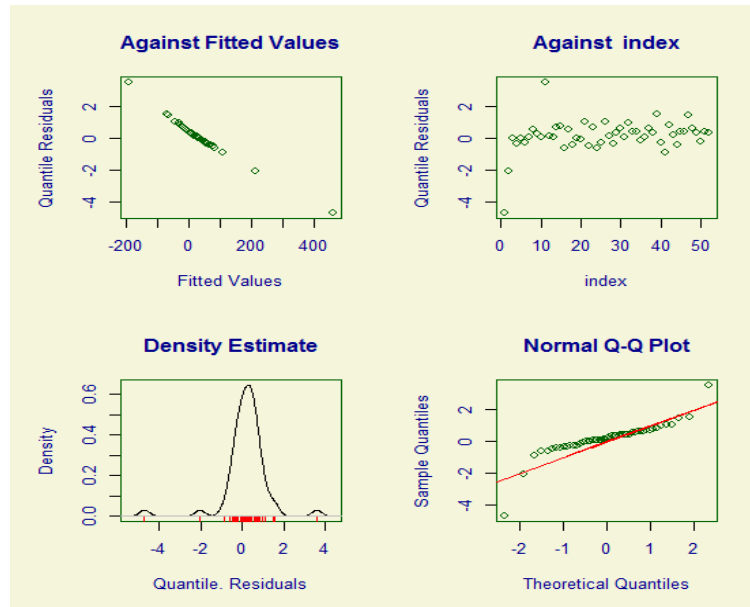
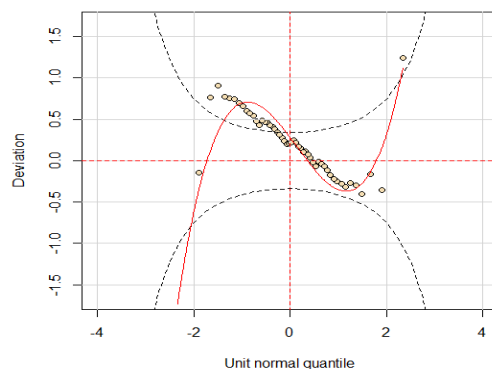


Figura 5-29: Prueba ajuste modelo Razonamiento Cuantitativo Factor Socioeconómico.

```
*****
Summary of the Quantile Residuals
      mean   = 0.1765264
      variance = 1.062699
      coef. of skewness = -1.476382
      coef. of kurtosis = 12.73828
Filliben correlation coefficient = 0.8693302
*****
```

En las figuras 5-28 y 5-29 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-30: Ajuste de la distribución modelo Razonamiento Cuantitativo Factor Socioeconómico.



A partir de la validación del modelo óptimo se puede decir que, el factor socioeconómico no influye en el promedio, la variación y el sesgo del puntaje de Razonamiento Cuantitativo de los estudiantes de Licenciatura en Matemáticas y Estadística.

El factor socioeconómico no es influyente en el puntaje de Razonamiento Cuantitativo que obtienen los estudiantes de Licenciatura en Matemáticas y Estadística en las pruebas Saber Pro, el puntaje se ve influenciado por variables del factor académico como son la educación del padre (Primaria y secundaria) y del factor Demográfico como el sexo, la edad, el número de personas en el hogar, el número de dormitorios, ser cabeza de familia y el tipo de situación del hogar.

5.3.2. Puntaje Inglés

A continuación se presentarán los modelos correspondientes a los factores Académicos, Demográficos y Socioeconómicos del puntaje de Inglés.

Modelo Factor Académico

La Figura 5-31 presenta el modelo óptimo correspondiente al Factor Académico.

Figura 5-31: Modelo óptimo Inglés Factor Académico.

```
*****
Family:  c("RG", "Reverse Gumbel")

Call:  gamlss(formula = INGLES_PUNT ~ EDUCAPADRE1, sigma.formula = ~1,
             family = RG, data = na.omit(Dataset), trace = FALSE)

Fitting method: RS()

-----
Mu link function:  identity
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	11.0000	0.5151	21.355	< 2e-16 ***
EDUCAPADRE1PRIMARIA	-1.5219	0.5250	-2.899	0.00563 **
EDUCAPADRE1SECUNDARIA	-1.4080	0.5264	-2.675	0.01019 *

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function:  log
Sigma Coefficients:

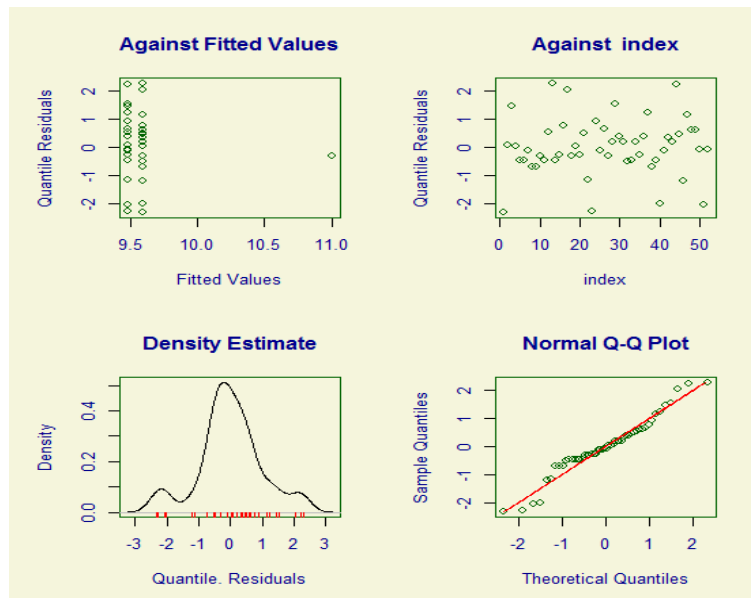
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.6634	0.1057	-6.275	9.53e-08 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  52
Degrees of Freedom for the fit:  4
    Residual Deg. of Freedom:  48
                        at cycle:  3

Global Deviance:  94.1337
      AIC:  102.1337
      SBC:  109.9387
*****
```

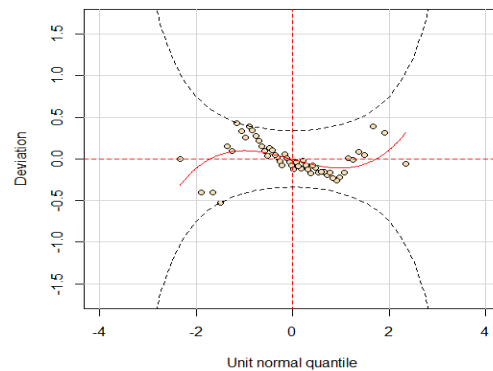
Figura 5-32: Diagnóstico de la distribución modelo Inglés Factor Académico.**Figura 5-33:** Prueba de ajuste modelo Inglés Factor Académico.

```

*****
Summary of the Quantile Residuals
      mean   = -0.004963944
      variance = 1.006917
      coef. of skewness = -0.05447635
      coef. of kurtosis  = 3.475612
Filliben correlation coefficient = 0.9787388
*****

```

En las figuras 5-32 y 5-33 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-34: Ajuste de la distribución modelo Inglés Factor Académico.

A partir de la validación del modelo óptimo se puede decir que, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo la educación del padre se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes donde el padre tiene educación primaria el puntaje promedio de Inglés disminuye 1.52 puntos respecto a los estudiantes con padres sin ningún tipo de estudios.

Con un nivel de significancia del 5 %, que aquellos estudiantes donde el padre tiene educación secundaria el puntaje promedio de Inglés disminuye 1.40 puntos respecto a los estudiantes con padres sin ningún tipo de estudios.

También se determina que la variabilidad del puntaje de Inglés de los estudiantes de Licenciatura en Matemáticas y Estadística no es explicada por el factor académico.

Modelo Factor Demográfico

Las Figuras 5-35 y 5-36 presentan el modelo óptimo correspondiente al Factor Demográfico.

Figura 5-35: Modelo óptimo Inglés Factor Demográfico.

```
*****
Family:  c("RG", "Reverse Gumbel")

Call:  gamlss(formula = INGLES_PUNT ~ ESTADO.CIVIL + CABEZAFMLIA + EDAD +
  SITUACIÓNHOGAR + NUMERO.DE.PERSOS.A.CARGO, sigma.formula = ~ESTADO.CIVIL +
  GÉNERO + CABEZAFMLIA + EDAD + SITUACIÓNHOGAR + NÚMERO.PERSOS.HOGAR +
  NUMERO.DE.PERSOS.A.CARGO, family = RG, data = na.omit(Dataset),      trace = FALSE)

Fitting method: RS()
```

Continuación del Modelo óptimo Inglés Factor Demográfico

Figura 5-36: Modelo óptimo Inglés Factor Demográfico.

```

fu link function: identity
fu Coefficients:

                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    9.204340   0.171705  53.605 < 2e-16 ***
ESTADO.CIVILcasado              0.865246   0.164779   5.251 3.97e-06 ***
ESTADO.CIVILSoltero            0.002135   0.166054   0.013 0.98980
EABEZAFMLIASi                  -0.562128   0.054518 -10.311 1.98e-13 ***
EDAD                           0.012362   0.002588   4.777 1.92e-05 ***
SITUACIÓNHOGAR{temporal por razones de estudio} 0.168758   0.061592   2.740 0.00878 **
NUMERO.DE.PERSOS.A.CARGO       -0.184474   0.029929  -6.164 1.78e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----

sigma link function: log
sigma Coefficients:

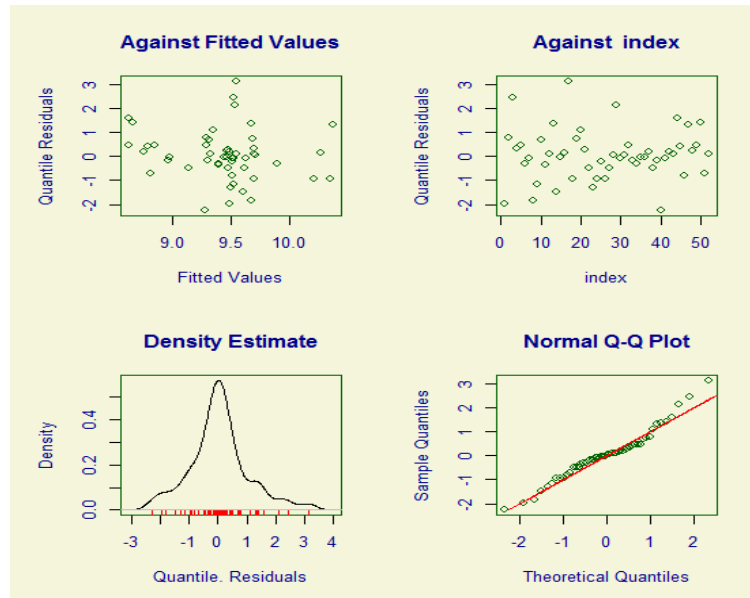
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    7.08690    0.95449   7.425 3.14e-09 ***
ESTADO.CIVILcasado            -1.67259    0.49128  -3.405 0.00145 **
ESTADO.CIVILSoltero          -0.41438    0.46933  -0.883 0.38219
GÉNERO                         -0.08925    0.23487  -0.380 0.70583
EABEZAFMLIASi                 0.79828    0.35695   2.236 0.03057 *
EDAD                          -0.23951    0.02816 -8.504 9.27e-11 ***
SITUACIÓNHOGAR{temporal por razones de estudio} -0.61874    0.24003  -2.578 0.01345 *
NÚMERO.PERSOS.HOGAR          -0.30136    0.06667  -4.520 4.79e-05 ***
NUMERO.DE.PERSOS.A.CARGO      0.06742    0.16732   0.403 0.68896
. |
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----

No. of observations in the fit: 52
Degrees of Freedom for the fit: 16
    Residual Deg. of Freedom: 36
                        at cycle: 8

Global Deviance:    55.05385
      AIC:          87.05385
      SBC:          118.2738
*****

```

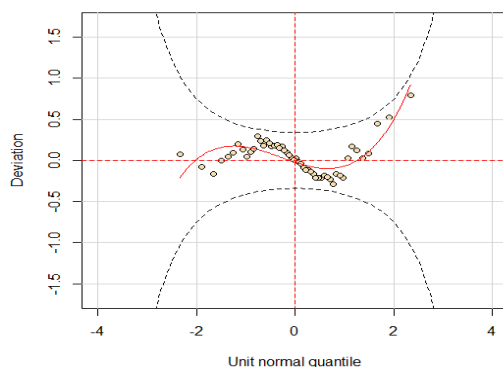
Figura 5-37: Diagnóstico de la distribución modelo Inglés Factor Demográfico.**Figura 5-38:** Prueba ajuste modelo Inglés Factor Demográfico.

```

*****
Summary of the Quantile Residuals
      mean   = 0.0462793
    variance = 1.071253
  coef. of skewness = 0.4326075
  coef. of kurtosis = 3.866087
Filliben correlation coefficient = 0.9799282
*****

```

En las figuras 5-37 y 5-38 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-39: Ajuste de la distribución modelo Inglés Factor Demográfico.**Figura 5-40:** Valores atípicos Número de personas a cargo Inglés.

```
> Q.stats(w,xvar=Dataset$NUMERO.DE.PERSOS.A.CARGO,n.inter=4)
```

	Z1	Z2	Z3	Z4	AgostinoK2	N
-0.5 to 0.5	-0.62905135	0.2360715	-2.47632577	0.6684711	6.57904293	55
0.5 to 1.5	1.45646080	-0.8547900	2.18448123	2.3121140	10.11782926	22
1.5 to 5.5	-0.72519241	-0.8496216	0.18370454	0.2058727	0.07613094	19
TOTAL Q stats	3.04288770	1.5082526	10.93789492	5.8351082	16.77300313	96
df for Q stats	1.00000000	2.00000000	3.00000000	3.00000000	6.00000000	0
p-val for Q stats	0.08109155	0.4704214	0.01206649	0.1199140	0.01015471	0

Figura 5-41: Valores atípicos Edad Inglés.

```
> Q.stats(w1,xvar=Dataset$EDAD,n.inter=4)
```

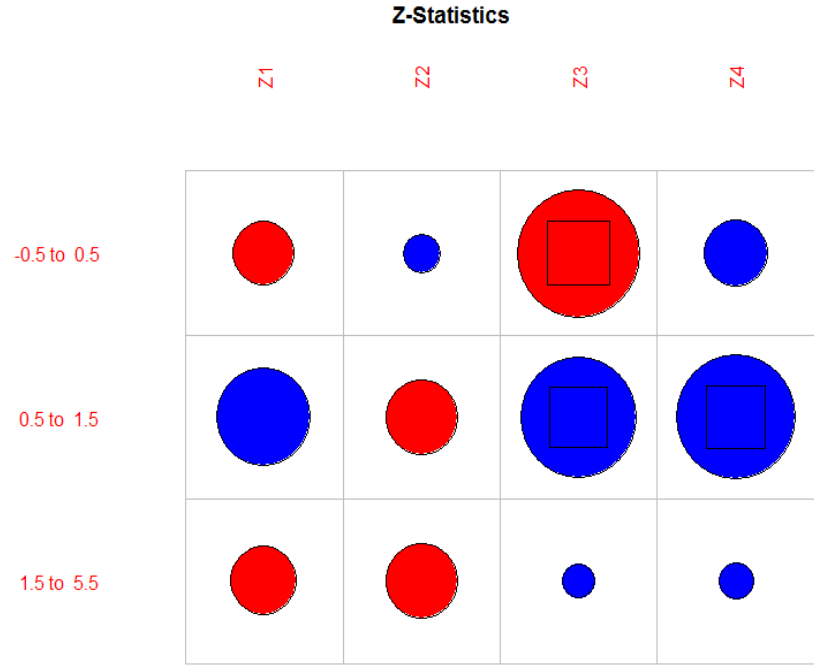
	Z1	Z2	Z3	Z4	AgostinoK2	N
19.5 to 24.5	0.7186429	0.9268229	-1.56509323	1.989918931	6.40929418	33
24.5 to 26.5	-0.9906376	-1.4149518	-1.64539937	-0.006579485	2.70738236	22
26.5 to 29.5	0.3923826	0.1338622	-2.27556251	1.774568972	8.32727975	18
29.5 to 51.5	-0.5067847	-1.3635553	0.27213514	-0.095023911	0.08308708	23
TOTAL Q stats	1.9086053	4.7382913	10.40909815	7.117945223	17.52704337	96
df for Q stats	2.00000000	3.00000000	4.00000000	4.00000000	8.00000000	0
p-val for Q stats	0.3850806	0.1919949	0.03407244	0.129785094	0.02506570	0

Figura 5-42: Valores atípicos Número personas hogar Inglés.

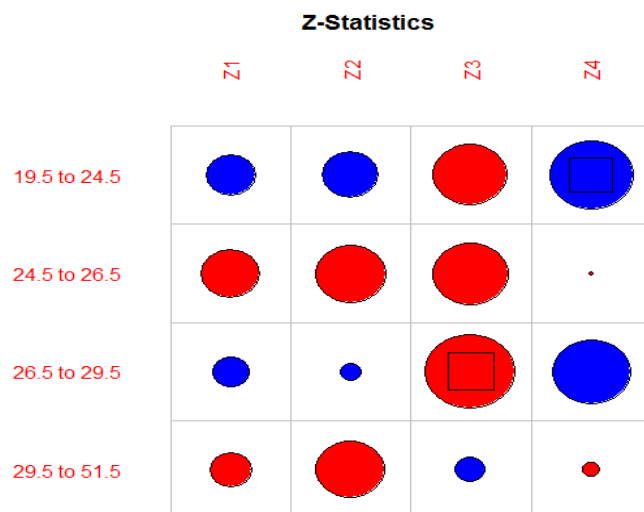
```
> Q.stats(w2,xvar=Dataset$NÚMERO.PERSOS.HOGAR ,n.inter=4)
```

	Z1	Z2	Z3	Z4	AgostinoK2	N
0.5 to 3.5	0.49312426	-1.37767281	-2.02177472	1.2529576	5.6574758	35
3.5 to 4.5	-1.64811453	-0.35563876	-1.69906018	1.0157539	3.9185615	23
4.5 to 6.5	0.07096107	1.17282737	0.07293541	1.0559838	1.1204214	30
6.5 to 9.5	1.23844917	-2.45538507	0.57720190	-0.3081785	0.4281360	8
TOTAL Q stats	4.49824488	9.42890118	7.31286013	3.8117346	11.1245948	96
df for Q stats	2.00000000	3.00000000	4.00000000	4.0000000	8.0000000	0
p-val for Q stats	0.10549176	0.02409989	0.12025017	0.4320839	0.1947389	0

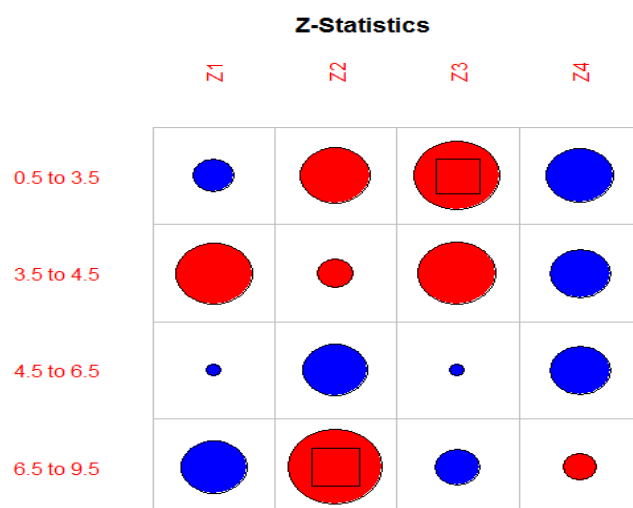
Figura 5-43: Prueba de valores atípicos número de personas a cargo Inglés.



Se presentan datos atípicos en la variable, indicando un bajo (-2.47) y alto (2.18 y 2.31) número de personas a cargo.

Figura 5-44: Prueba de valores atípicos Edad Inglés.

Se presentan datos atípicos en la variable, indicando un bajo número de años (-2.27) y no significativo (1.98).

Figura 5-45: Prueba de valores atípicos Número de personas en el hogar Inglés.

Se presentan datos atípicos en la variable, indicando un bajo número de personas a cargo (-2.45) y (-2.0).

A partir de la validación del modelo óptimo se puede decir que, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo su estado civil se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes casados el puntaje promedio de Inglés aumenta 0.86 puntos respecto a los estudiantes en unión libre.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el hecho de ser cabeza de familia se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que son cabeza de familia, en promedio disminuye 0.56 puntos en el puntaje de Inglés.

Al comparar dos estudiantes donde el segundo difiere del primero en un año de más en su edad se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que presenta mayor edad, obtienen en promedio un aumento de 0.012 puntos en el puntaje de Inglés.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo su situación del hogar se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes con hogar temporal por razones de estudio, obtienen en promedio un aumento de 0.17 puntos respecto a los estudiantes con hogar habitual o permanente.

Al comparar dos estudiantes donde el segundo difiere del primero en una persona de más a cargo se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que tienen personas a cargo, en promedio disminuyen 0.18 puntos en el puntaje de Inglés.

También se determina que al comparar dos estudiantes donde el segundo difiere del primero en tan sólo su estado civil se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes casados, obtienen una disminución de la variabilidad en el puntaje de Inglés respecto a los estudiantes en unión libre.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el hecho de ser cabeza de familia se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que son cabeza de familia, obtienen una disminución de la variabilidad en el puntaje de Inglés respecto a

los estudiantes que no lo son.

Al comparar dos estudiantes donde el segundo difiere del primero en un año de más en su edad se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que presentan mayor edad, disminuyen la variabilidad en el puntaje de Inglés.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo su situación del hogar se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes con hogar temporal por razones de estudio, obtienen una disminución en su variabilidad respecto a los estudiantes con hogar habitual o permanente.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo una persona de más en el hogar se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes con una persona de más en el hogar, la variabilidad disminuye respecto a los que no.

Modelo Factor Socioeconómico

Las Figuras 5-46 y 5-47 presentan el modelo óptimo correspondiente al Factor Socioeconómico.

Figura 5-46: Modelo óptimo Inglés Factor Socioeconómico.

```
*****
Family: c("RG", "Reverse Gumbel")

Call: gamlss(formula = INGLES_PUNT ~ OCUP_MADRE1 + ESTRATORECIBOENERGIA + VALR_MATRICULA_AÑO.ANTERIOR + CELULAR + TV + TELEFONIA +
  COMPUTADOR + LAVADORA + NÚMERODORMITORIOS + HORAS_TRABAJO + TRABAJA, sigma.formula = ~1, family = RG, data = na.omit(Dataset), trace =

Fitting method: RS()
```

Continuación Modelo óptimo Inglés Factor Socioeconómico

Figura 5-47: Modelo óptimo Inglés Factor Socioeconómico.

```

-----
Mu link function: identity
Mu Coefficients:

                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   11.91274    0.71186   16.735 < 2e-16 ***
OCUP_MADRE1Obrero              -1.11371    0.36151   -3.081  0.00422 **
OCUP_MADRE1Ocupación no remunerado -0.18393    0.18356   -1.002  0.32385
ESTRATORECIBOENERGIAESTR2      -0.37761    0.20656   -1.828  0.07687 .
ESTRATORECIBOENERGIAESTR3        0.36410    0.22359    1.628  0.11325
ESTRATORECIBOENERGIAESTR4        1.31285    0.49117    2.673  0.01174 *
VALR_MATRICULA_AÑO.ANTERIOR{Entre 3 millones y 5 millones\r\r\n} -1.98730    0.60965   -3.260  0.00265 **
VALR_MATRICULA_AÑO.ANTERIOR{Entre 500 mil y menos de 1 millón} -0.05994    0.34940   -0.172  0.86488
VALR_MATRICULA_AÑO.ANTERIOR{Más de 5 millones\r\r\n} -0.24245    0.52343   -0.463  0.64635
VALR_MATRICULA_AÑO.ANTERIOR{Menos de 500 mil\r\r\n} -0.62968    0.35318   -1.783  0.08409 .
CELULARsi                      -0.98206    0.32880   -2.987  0.00537 **
TVsi                           0.48243    0.17240    2.798  0.00863 **
TELEFONIASi                    0.37112    0.21372    1.736  0.09210 .
COMPUTADORsi                   0.75797    0.23246    3.261  0.00264 **
LAVADORAsi                     -0.49745    0.20780   -2.394  0.02271 *
NÚMERODORMITORIOS              -0.20424    0.07609   -2.684  0.01142 *
HORAS_TRABAJO                  -0.03490    0.01569   -2.225  0.03327 *
TRABAJA{si, para tener experiencia} -0.50731    0.22309   -2.274  0.02981 *
TRABAJA{si, por ser plan de practica}  0.60009    0.47551    1.262  0.21607
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

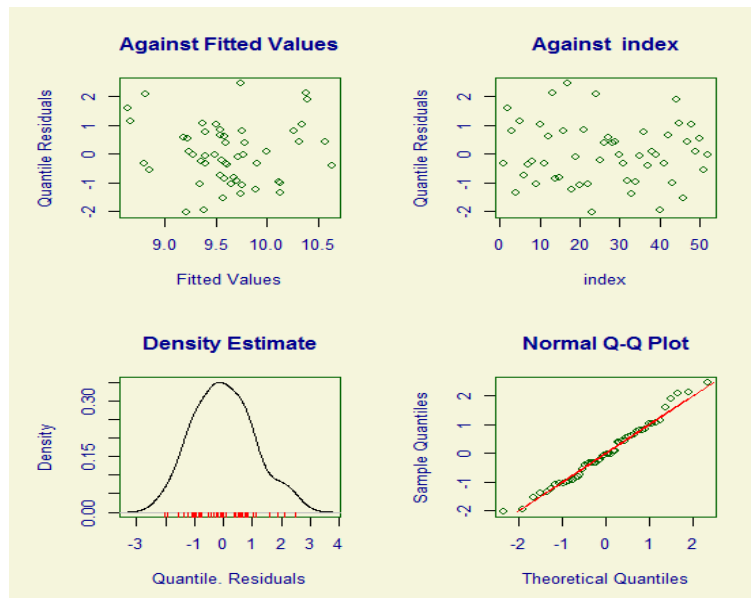
-----
Sigma link function: log
Sigma Coefficients:

                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -0.9825     0.1146   -8.57 8.57e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit:  52
Degrees of Freedom for the fit:  20
    Residual Deg. of Freedom:  32
                        at cycle:  4

Global Deviance:    64.7694
      AIC:          104.7694
      SBC:          143.7943
*****

```

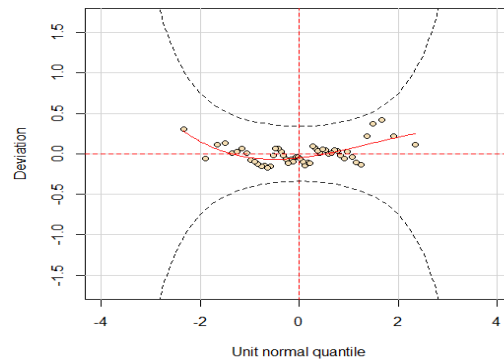
Figura 5-48: Diagnóstico de la distribución modelo Inglés Factor Socioeconómico.**Figura 5-49:** Prueba ajuste modelo Inglés Factor Socioeconómico.

```

*****
Summary of the Quantile Residuals
      mean   = 0.005822305
      variance = 1.073508
      coef. of skewness = 0.2781637
      coef. of kurtosis = 2.594026
Filliben correlation coefficient = 0.9926096
*****

```

En las figuras 5-48 y 5-49 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-50: Ajuste de la distribución modelo Inglés Factor Socioeconómico.

A partir de la validación del modelo óptimo se puede decir que, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo la ocupación de la madre se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes donde la ocupación de la madre es obrero el puntaje promedio de Inglés disminuye 1.11 puntos respecto a los estudiantes con ocupación de la madre de administrador.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el estrato del recibo de energía se tiene que, con un nivel de significancia del 10 %, que aquellos estudiantes presentan un estrato dos, el puntaje promedio de Inglés disminuye 0.37 puntos respecto a los estudiantes de estrato uno. Con un nivel de significancia del 5 %, que aquellos estudiantes presentan un estrato cuatro, el puntaje promedio de Inglés aumenta 1.31 puntos respecto a los estudiantes de estrato uno.

Al comparar dos estudiantes donde el segundo difiere del primero en tan el valor de la matrícula del semestre anterior se tiene que, con un nivel de significancia inferior del 5 %, que aquellos estudiantes presentan un valor de la matrícula del semestre anterior entre tres y cinco smmlv, el puntaje promedio de Inglés disminuye 1.98 puntos respecto a los estudiantes con un salario familiar mensual menor a dos y tres smmlv.. Con un nivel de significancia del 10 %, que aquellos estudiantes presentan un valor de la matrícula del semestre anterior de 500 mil, el puntaje promedio de Inglés disminuye 0.62 puntos respecto a los estudiantes con un salario familiar mensual menor a dos y tres smmlv..

Al comparar dos estudiantes donde el segundo difiere del primero en tan solo el hecho de tener celular se tiene que, con un nivel de significancia inferior del 5 %, que aquellos

estudiantes con celular, el puntaje promedio de Inglés disminuye 0.98 puntos respecto a los estudiantes que no tienen.

Al comparar dos estudiantes donde el segundo difiere del primero en tan solo el hecho de tener servicio de televisión se tiene que, con un nivel de significancia del 10 %, que aquellos estudiantes con servicio de televisión, el puntaje promedio de Inglés aumenta 0.48 puntos respecto a los estudiantes que no tienen.

Al comparar dos estudiantes donde el segundo difiere del primero en tan solo el hecho de tener servicio de telefonía se tiene que, con un nivel de significancia del 10 %, que aquellos estudiantes con servicio de telefonía, el puntaje promedio de Inglés aumenta 0.37 puntos respecto a los estudiantes que no tienen.

Al comparar dos estudiantes donde el segundo difiere del primero en tan solo el hecho de tener computador se tiene que, con un nivel de significancia inferior del 5 %, que aquellos estudiantes con computador, el puntaje promedio de Inglés disminuye 0.75 puntos respecto a los estudiantes que no tienen.

Al comparar dos estudiantes donde el segundo difiere del primero en tan solo el hecho de tener lavadora se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes con lavadora, el puntaje promedio de Inglés disminuye 0.49 puntos respecto a los estudiantes que no tienen.

Al comparar dos estudiantes donde el segundo difiere del primero en tan solo un dormitorio de más en el hogar se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes con más dormitorios en el hogar, el puntaje promedio de Inglés disminuye 0.20 puntos.

Al comparar dos estudiantes donde el segundo difiere del primero en tan solo una hora de más en el trabajo se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes con más horas de trabajo, el puntaje promedio de Inglés disminuye 0.034 puntos.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo en el tipo de trabajo se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes

con un trabajo por tener experiencia, el puntaje promedio de Inglés disminuye 0.50 puntos respecto a los estudiantes de que no trabajan.

También se determina que la variabilidad del puntaje de Inglés de los estudiantes de Licenciatura en Matemáticas y Estadística no se explica por el factor socioeconómico.

Los factores Académico y Socioeconómico no son incidentes en la variabilidad del puntaje obtenido en Inglés por los estudiantes de Licenciatura en Matemáticas y Estadística, por el contrario el factor Demográfico presenta en su mayoría de variables como Valor de la matrícula del semestre anterior, ocupación de la madre, estrato de recibo de energía, celular, telefonía, televisión, lavadora, número de dormitorios, horas de trabajo y trabaja, una incidencia en este puntaje.

5.3.3. Puntaje Lectura Crítica

A continuación se presentarán los modelos correspondientes a los factores Académicos, Demográficos y Socioeconómicos del puntaje de Lectura Crítica.

Modelo Factor Académico

Las Figuras 5-51 y 5-52 presentan el modelo óptimo correspondiente al Factor Académico.

Figura 5-51: Modelo óptimo Lectura Crítica Factor Académico.

```
Call:  gamlss(formula = LECTURA_CRITICA ~ EDUCAMADRE1, sigma.formula =
nu.formula = ~1, family = exGAUS, data = na.omit(Dataset),
Fitting method: RS())

-----
Mu link function:  identity
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.7473      0.2396  40.682  <2e-16 ***
EDUCAMADRE1Primaria -0.5039      0.2549  -1.977   0.0538 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function:  log
Sigma Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.3825      0.1595  -2.399   0.0204 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Continuación Modelo óptimo Lectura Crítica Factor Académico

Figura 5-52: Modelo óptimo Lectura Crítica Factor Académico.

```
-----
Nu link function: log
Nu Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.3255      0.2511  -1.296   0.201

-----

No. of observations in the fit: 52
Degrees of Freedom for the fit:  4
Residual Deg. of Freedom: 48
      at cycle: 10

Global Deviance:    143.5932
      AIC:           151.5932
      SBC:           159.3982
*****
```

Figura 5-53: Diagnóstico de la distribución modelo Lectura Crítica Factor Académico.

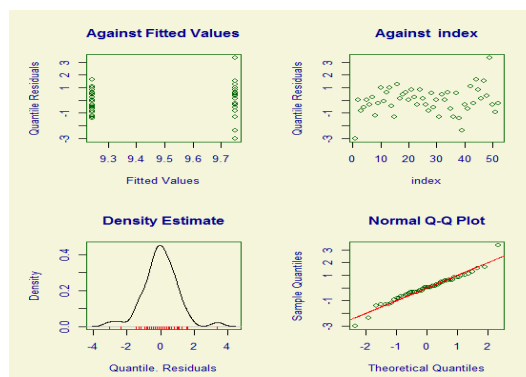
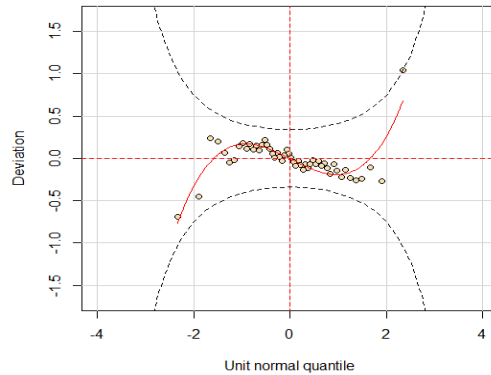


Figura 5-54: Prueba ajuste modelo Lectura Crítica Factor Académico.

```
*****
Summary of the Quantile Residuals
      mean    = -0.01095265
      variance = 1.051483
      coef. of skewness = 0.04291727
      coef. of kurtosis = 4.939395
Filliben correlation coefficient = 0.9752785
*****
```

En las figuras 5-53 y 5-54 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-55: Ajuste de la distribución modelo Lectura Crítica Factor Académico.

A partir de la validación del modelo óptimo se puede decir que, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo en la educación de la madre se tiene que, con un nivel de significancia del 10 %, que aquellos estudiantes donde la madre tiene educación primaria, el puntaje promedio de Lectura Crítica disminuye 0.50 puntos respecto a los estudiantes con madre sin ningún tipo de estudios.

También se determina que la variabilidad y el sesgo del puntaje de Lectura Crítica de los estudiantes de Licenciatura en Matemáticas y Estadística no son explicados por el factor académico.

Modelo Factor Demográfico

Las Figuras 5-56 y 5-57 presentan el modelo óptimo correspondiente al Factor Demográfico.

Figura 5-56: Modelo óptimo Lectura Crítica Factor Demográfico.

```
*****
Family:  c("exGAUS", "ex-Gaussian")

Call:  gamlss(formula = LECTURA_CRITICA ~ CABEZAFMLIA + NÚMERO.PERSOS.HOGAR +
  NUMERO.DE.PERSOS.A.CARGO, sigma.formula = ~1, nu.formula = ~1,
  family = exGAUS, data = na.omit(Dataset), trace = FALSE)

Fitting method: RS()
```

Continuación Modelo óptimo Lectura Crítica Factor Demográfico.

Figura 5-57: Modelo óptimo Lectura Crítica Factor Demográfico.

```
-----
Mu link function:  identity
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.95283    0.32388  30.730  <2e-16 ***
CABEZAFMLIASi  -0.93358    0.34225  -2.728   0.0090 **
NÚMERO.PERSOS.HOGAR -0.16910    0.07438  -2.274   0.0277 *
NUMERO.DE.PERSOS.A.CARGO 0.33229    0.14326   2.320   0.0249 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----

Sigma link function:  log
Sigma Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.7083    0.2681  -2.642   0.0112 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----

Nu link function:  log
Nu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.09369    0.22730  -0.412   0.682

-----

No. of observations in the fit:  52
Degrees of Freedom for the fit:  6
Residual Deg. of Freedom:  46
Global Deviance:    139.5579
AIC:    151.5579
SBC:    163.2654
*****
```

Figura 5-58: Diagnóstico de la distribución modelo Lectura Crítica Factor Demográfico.

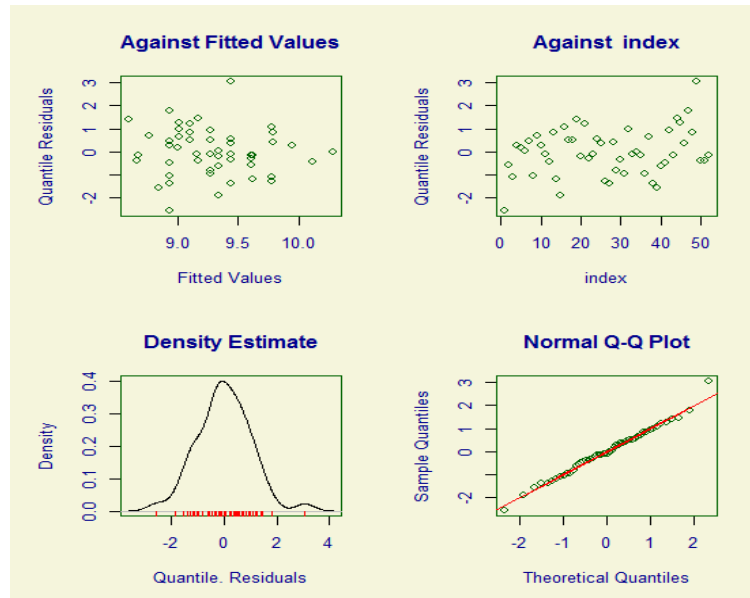


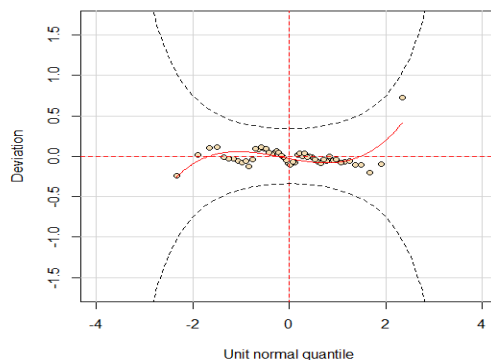
Figura 5-59: Prueba ajuste modelo Lectura Crítica- Factor Demográfico.

```

*****
Summary of the Quantile Residuals
      mean   = -0.008387572
      variance = 1.034948
      coef. of skewness = 0.1631325
      coef. of kurtosis = 3.536465
Filliben correlation coefficient = 0.9920165
*****

```

En las figuras 5-58 y 5-59 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-60: Ajuste de la distribución modelo Lectura Crítica Factor Demográfico.**Figura 5-61:** Valores Atípicos Número de personas a cargo Lectura Crítica.

```
> Q.stats(w,xvar=Dataset$NUMERO.DE.PERSOS.A.CARGO,n.inter=4)
```

	Z1	Z2	Z3	Z4	AgostinoK2	N
-0.5 to 0.5	-0.8354622	0.3172633	-1.225159621	-0.91481995	2.337911647	55
0.5 to 1.5	1.5473493	0.0183220	2.583642441	2.49722638	12.911347855	22
1.5 to 5.5	-0.3515347	-0.6382849	1.024406809	0.07258482	1.054677867	19
TOTAL Q stats	3.2158635	0.5083993	9.225633669	7.07830370	16.303937369	96
df for Q stats	1.0000000	2.0000000	2.000000000	3.00000000	5.000000000	0
p-val for Q stats	0.0729277	0.7755370	0.009923825	0.06944341	0.006027824	0

Figura 5-62: valores atípicos Edad Lectura Crítica.

```
> Q.stats(w1,xvar=Dataset$EDAD,n.inter=4)
```

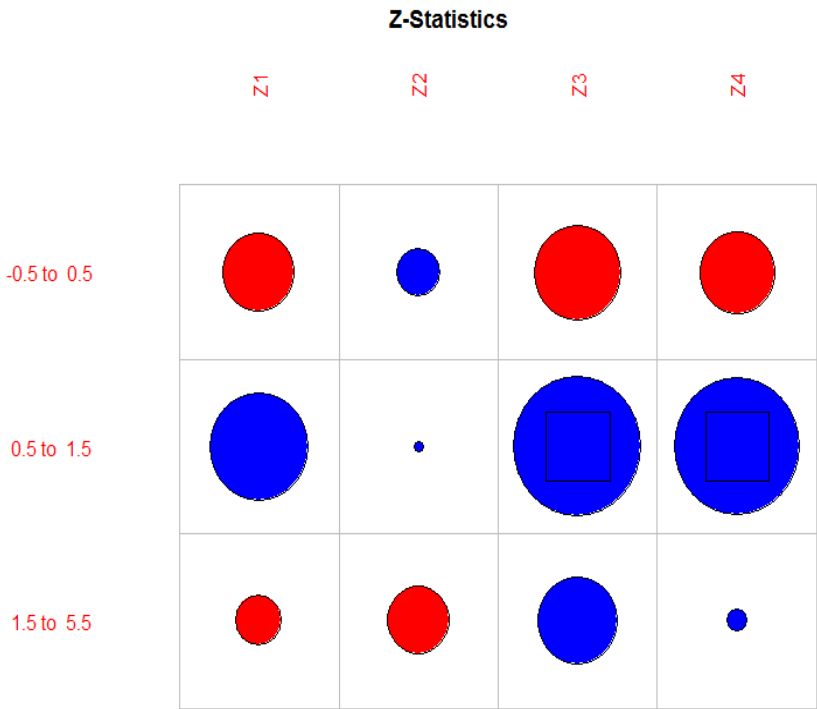
	Z1	Z2	Z3	Z4	AgostinoK2	N
19.5 to 24.5	0.6788505	1.3667427	0.3041232	1.71374838	3.0294244	33
24.5 to 26.5	-1.2486127	-1.3237703	-1.0917818	-0.54472873	1.4887169	22
26.5 to 29.5	0.6007930	0.3086510	-1.2394300	0.13177133	1.5535505	18
29.5 to 51.5	-0.2199836	-0.8388824	0.9191631	-0.07111639	0.8499183	23
TOTAL Q stats	2.4292166	4.4193426	3.6655260	3.25608412	6.9216101	96
df for Q stats	2.0000000	3.0000000	3.0000000	4.00000000	7.0000000	0
p-val for Q stats	0.2968263	0.2195986	0.2999199	0.51592031	0.4370872	0

Figura 5-63: valores atípicos Número personas hogar Lectura Crítica.

```
> Q.stats(w2,xvar=Dataset$NUMERO.PERSOS.HOGAR ,n.inter=4)
```

	Z1	Z2	Z3	Z4	AgostinoK2	N
0.5 to 3.5	0.56697684	-0.96730551	-0.8661441	0.03444588	0.7513921	35
3.5 to 4.5	-1.70503326	-0.53255341	-0.6112009	-0.45549470	0.5810419	23
4.5 to 6.5	0.18559325	1.72970237	1.2742430	1.12477996	2.8888251	30
6.5 to 9.5	1.17592350	-2.05312144	0.6657799	-0.29121272	0.5280678	8
TOTAL Q stats	4.64584210	8.42647105	3.1907302	1.55859675	4.7493270	96
df for Q stats	2.00000000	3.00000000	3.00000000	4.00000000	7.00000000	0
p-val for Q stats	0.09798694	0.03797301	0.3631428	0.81621355	0.6905229	0

Figura 5-64: Prueba de valores atípicos número de personas a cargo Lectura Crítica.



La variable presenta valores atípicos, indicando un alto número de personas a cargo (2.49 y 2.58).

Figura 5-65: Prueba de valores Atípicos Edad Lectura Crítica.

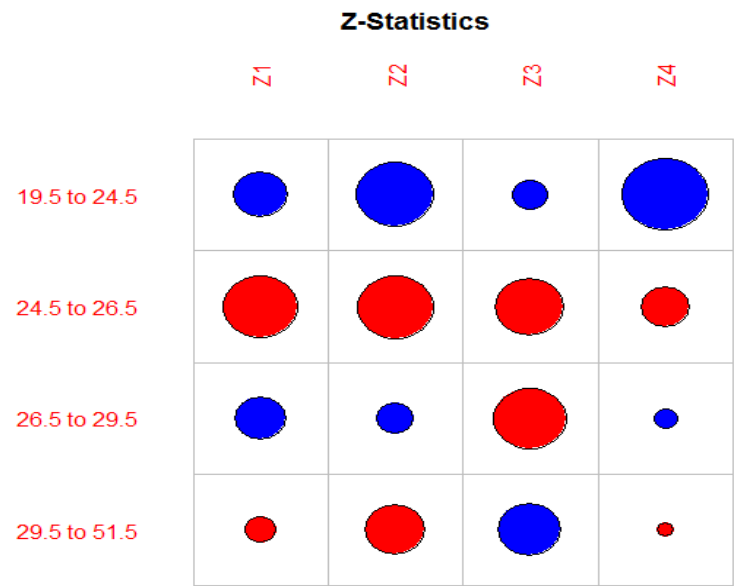
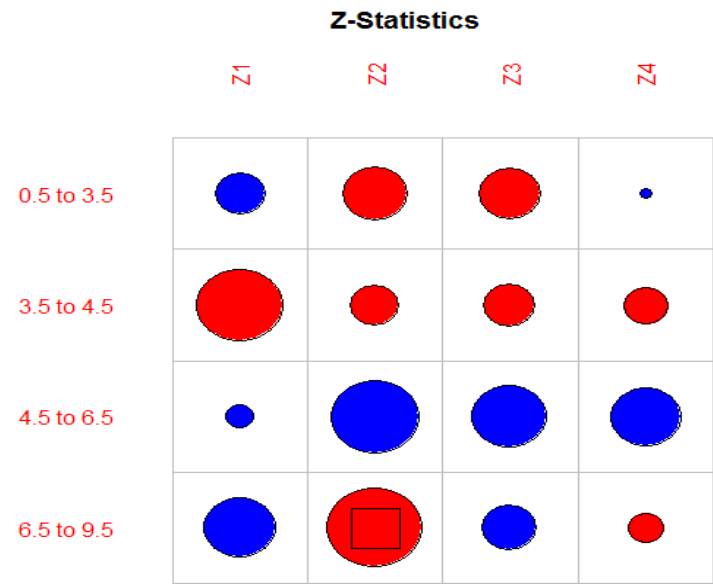


Figura 5-66: Prueba de valores atípicos Número de personas en el hogar Lectura Crítica.



La variable edad no presenta datos típicos, mientras que número de personas en el hogar si, indicando un bajo número de personal en el hogar (-2.05).

A partir de la validación del modelo óptimo se puede decir que, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el hecho de ser cabeza de familia se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes cabeza de familia, el puntaje promedio de Lectura Crítica disminuye 0.53 puntos respecto a los estudiantes que no lo son.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo una persona de más en el hogar, se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que presentan un mayor número de personas en el hogar, obtienen en promedio una disminución de 0.17 puntos en el puntaje de Lectura Crítica.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el número de personas a cargo, se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que presentan un mayor número de personas en el hogar, obtienen en promedio una aumento de 0.32 puntos en el puntaje de Lectura Crítica.

También se determino la variabilidad y el sesgo del puntaje de Lectura Crítica de los estudiantes de Licenciatura en Matemáticas y Estadística no son explicados por el factor demográfico.

Modelo Factor Socioeconómico

Las Figuras 5-67 y 5-68 presenta el modelo óptimo correspondiente al Factor Socioeconómico.

Figura 5-67: Modelo óptimo Lectura Crítica Factor Socioeconómico.

```
*****  
Family: c("exGAUS", "ex-Gaussian")  
  
Call: gamlss(formula = LECTURA_CRITICA ~ INTERNET + TV + TELEFONIA + COMPUTADOR + LAVADORA + FMIAR_MENSUAL + NÚMERODORMITORIOS +  
      HORAS_TRABAJO + TRABAJO, sigma.formula = ~1, nu.formula = ~1, family = exGAUS, data = na.omit(Dataset))  
  
Fitting method: RS()
```

Continuación Modelo óptimo Lectura Crítica Factor Socioeconómico.

Figura 5-68: Modelo óptimo Lectura Crítica Factor Socioeconómico.

```
-----
Mu link function: identity
Mu Coefficients:

                Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.95674    0.72353  12.379 4.37e-15 ***
INTERNETsi       0.14328    0.32895   0.436 0.66554
TVsi             0.90956    0.31374   2.899 0.00611 **
TELEFONIASi      0.15177    0.35534   0.427 0.67165
COMPUTADORsi     0.58375    0.45538   1.282 0.20745
LAVADORAsi      -0.41258    0.42412  -0.973 0.33665
FMIAR_MENSUAL(3 y 4 menos sm) 0.30070    0.48860   0.615 0.54184
FMIAR_MENSUAL(4 y 5 menos sm) 2.88195    0.46410   6.210 2.65e-07 ***
FMIAR_MENSUALism -0.03174    0.36660  -0.087 0.93146
NÚMERODORMITORIOS 0.09377    0.17298   0.542 0.59083
HORAS_TRABAJO    -0.02687    0.02265  -1.186 0.24267
TRABAJA(si, para tener experiencia) -0.09228    0.32215  -0.286 0.77604
TRABAJA(si, por ser plan de practica) -2.33946    3.08889  -0.757 0.45338
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function: log
Sigma Coefficients:

                Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.009291    0.042750  -0.217 0.829

-----
Nu link function: log
Nu Coefficients:

                Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.4558     0.1752  -2.601 0.0121 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 52
Degrees of Freedom for the fit: 15
Residual Deg. of Freedom: 37
                        at cycle: 20

Global Deviance: 166.955
AIC: 196.955
SBC: 226.2236
*****
```

Figura 5-69: Diagnóstico de la distribución modelo Lectura Crítica Factor Socio-económico.

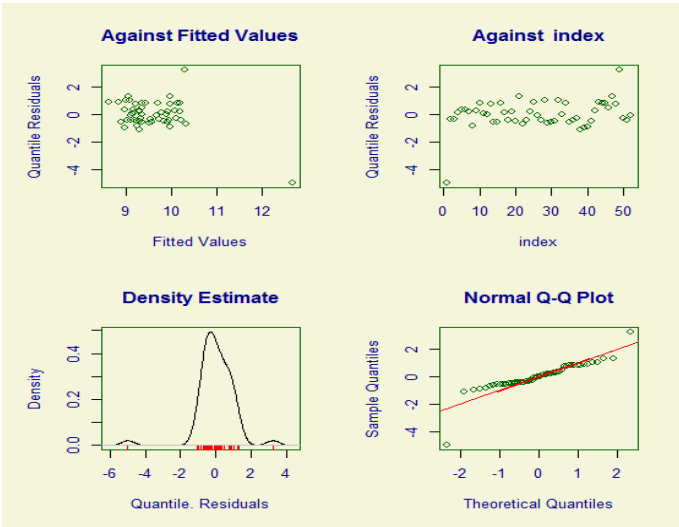
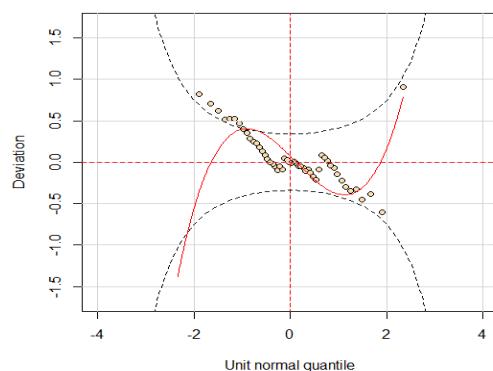


Figura 5-70: Prueba ajuste modelo Lectura Crítica Factor Socioeconómico.

```
*****
Summary of the Quantile Residuals
      mean   = 0.008844209
      variance = 1.091687
      coef. of skewness = -1.448353
      coef. of kurtosis = 12.20805
Filliben correlation coefficient = 0.8854918
*****
```

En las figuras 5-69 y 5-70 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-71: Ajuste de la distribución modelo Lectura Crítica Factor Socioeconómico.

A partir de la validación del modelo óptimo se puede decir que, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el hecho tener servicio de televisión se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes con servicio de televisión, el puntaje promedio de Lectura Crítica aumenta 0.90 puntos respecto a los estudiantes que no cuentan con este servicio.

Al comparar dos estudiantes donde el segundo difiere del primero en los ingresos familiares mensuales se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes con un salario familiar mensual menor a cuatro y cinco smmlv, el puntaje promedio de Lectura Crítica aumenta 2.88 puntos respecto a los estudiantes donde el ingreso familiar mensual es menor a dos y tres smmlv.

También se determino que la variabilidad y el sesgo del puntaje de Lectura Crítica no son explicados por los factores Académicos, Demográficos y Socioeconómicos, causado por problemas de convergencia en un modelo complejo.

5.3.4. Puntaje Comunicación Escrita

A continuación se presentarán los modelos correspondientes a los factores Académicos, Demográficos y Socioeconómicos del puntaje de Comunicación Escrita.

Modelo Factor Académico

Las Figura 5-72 presenta el modelo óptimo correspondiente al Factor Académico.

Figura 5-72: Modelo óptimo Comunicación Escrita Factor Académico.

```
*****
Family:  c("SN2", "skew normal type 2")

Call:  gamlss(formula = COMUNICA_ESCRITA_PUNT ~ EDUCAMADRE1 + EDUCAPADRE1 +
  NOCURSO, sigma.formula = ~EDUCAPADRE1 + NOCURSO, nu.formula = ~TIPOBACHILERRAT +
  EDUCAMADRE1 + EDUCAPADRE1 + NOCURSO, family = SN2, data = na.omit(Dataset),
  trace = FALSE)

Fitting method: RS()

-----
Mu link function:  identity
Mu Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)      79.2644    22.5561   3.514 0.000988 ***
EDUCAMADRE1Primaria  -1.0719     0.1990  -5.387 2.25e-06 ***
EDUCAPADRE1PRIMARIA  -67.7749    22.5548  -3.005 0.004251 **
EDUCAPADRE1SECUNDARIA -68.1968    22.5549  -3.024 0.004038 **
NOCURSORepaso por cuenta propia -0.1578     0.2025  -0.779 0.439820
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function:  log
Sigma Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.0265     6.1446  -0.167  0.868
EDUCAPADRE1PRIMARIA   0.7643     6.1417   0.124  0.901
EDUCAPADRE1SECUNDARIA  0.1493     6.1412   0.024  0.981
NOCURSORepaso por cuenta propia -0.1160     0.3013  -0.385  0.702

-----
Nu link function:  log
Nu Coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.41250    4.14418   0.582  0.56338
TIPOBACHILERRATN   0.07496    1.06064   0.071  0.94397
TIPOBACHILERRATT   0.35230    0.27886   1.263  0.21297
EDUCAMADRE1Primaria  0.93528    0.28852   3.242  0.00224 **
EDUCAPADRE1PRIMARIA -3.85175    4.13031  -0.933  0.35603
EDUCAPADRE1SECUNDARIA -3.85530    4.13436  -0.933  0.35605
NOCURSORepaso por cuenta propia 0.02622    0.27681   0.095  0.92496
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 5-73: Diagnóstico de la distribución modelo Comunicación Escrita Factor Académico.

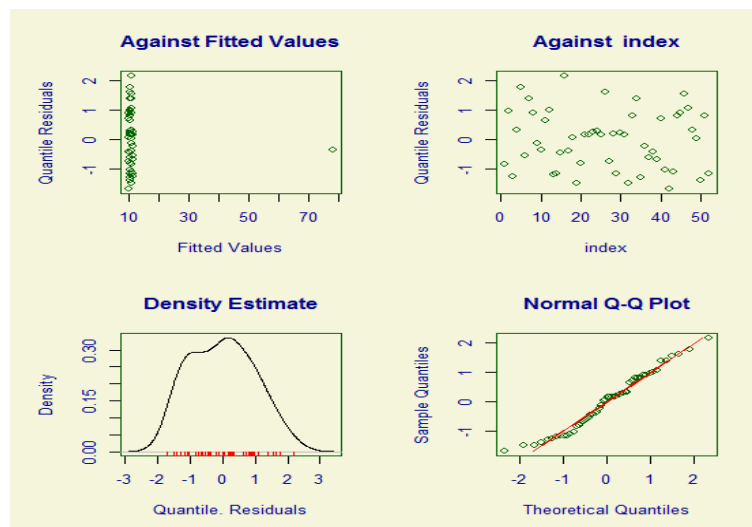
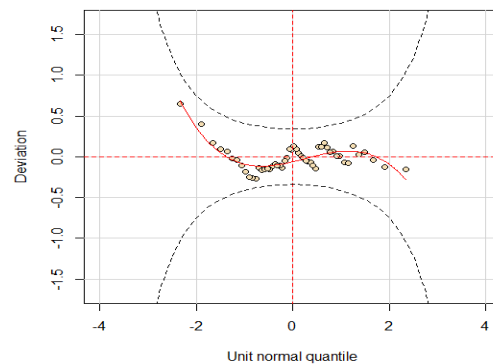


Figura 5-74: Prueba ajuste modelo Comunicación Escrita Factor Académico.

```
*****
Summary of the Quantile Residuals
      mean   = -0.01283238
    variance = 0.9639241
  coef. of skewness = 0.172392
  coef. of kurtosis = 2.046042
Filliben correlation coefficient = 0.9872774
*****
```

En las figuras 5-73 y 5-74 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-75: Ajuste de la distribución modelo Comunicación Escrita Factor Académico.

A partir de la validación del modelo óptimo se puede decir que, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo en la educación de la madre se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes donde la madre tiene educación primaria, el puntaje promedio de Comunicación Escrita disminuye 1.07 puntos respecto a los estudiantes con madre sin ningún tipo de estudios.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo en la educación del padre se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes donde el padre tiene educación primaria, el puntaje promedio de Comunicación Escrita disminuye respecto a los estudiantes con padre sin ningún tipo de estudios.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo en la educación del padre se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes donde el padre tiene educación secundaria, el puntaje promedio de Comunicación Escrita disminuye respecto a los estudiantes con padre sin ningún tipo de estudios.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo en la educación de la madre se tiene que, con un nivel de significancia inferior al 5 %, que la explicación del sesgo en el puntaje de Comunicación Escrita se da por la educación primaria de la madre respecto a los estudiantes con madre sin ningún tipo de estudios.

También se determinó que no existen variables del factor académico incidentes en la variabilidad del puntaje de los estudiantes Licenciatura en Matemáticas y Estadística

en Comunicación Escrita.

Modelo Factor Demográfico

La figura 5-76 presenta el modelo óptimo correspondiente al factor demográfico.

Figura 5-76: Modelo óptimo Comunicación Escrita Factor Demográfico.

```
*****
Family: c("SN2", "skew normal type 2")

Call: gamlss(formula = COMUNICA_ESCRITA_PUNT ~ ESTADO.CIVIL + CABEZAFMLIA + EDAD + NÚMERO.PERSOS.HOGAR + NUMERO.DE.PERSOS.A.CARGO,
  sigma.formula = ~CABEZAFMLIA + EDAD + SITUACIÓNHOGAR + NÚMERO.PERSOS.HOGAR + NUMERO.DE.PERSOS.A.CARGO,
  nu.formula = ~ESTADO.CIVIL + GÉNERO + CABEZAFMLIA + EDAD + SITUACIÓNHOGAR + NÚMERO.PERSOS.HOGAR +
    NUMERO.DE.PERSOS.A.CARGO, family = SN2, data = na.omit(Dataset), trace = FALSE)

Fitting method: RS()

-----
Mu link function: identity
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    17.1150812  0.1278985  133.818 < 2e-16 ***
ESTADO.CIVILcasado -1.9726561  0.0940569  -20.973 < 2e-16 ***
ESTADO.CIVILSoltero -2.6214396  0.0986483  -26.574 < 2e-16 ***
CABEZAFMLIAi -1.8904876  0.0737619  -25.630 < 2e-16 ***
EDAD           -0.1410615  0.0002731  -516.583 < 2e-16 ***
NÚMERO.PERSOS.HOGAR -0.2025568  0.0211576   -9.574 1.25e-10 ***
NUMERO.DE.PERSOS.A.CARGO 0.3827058  0.0660103    5.798 2.46e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function: log
Sigma Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -2.36098  0.91006  -2.594 0.01452 *
CABEZAFMLIAi    0.44457  0.65514   0.679 0.50260
EDAD            0.05075  0.03415   1.486 0.14769
SITUACIÓNHOGAR{temporal por razones de estudio} 0.42074  0.33056   1.273 0.21287
NÚMERO.PERSOS.HOGAR 0.10343  0.09113   1.135 0.26534
NUMERO.DE.PERSOS.A.CARGO -0.79115  0.22338  -3.542 0.00132 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function: log
Nu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -4.69436  0.70096  -6.697 2.02e-07 ***
ESTADO.CIVILcasado 2.98499  0.52917   5.641 3.82e-06 ***
ESTADO.CIVILSoltero 2.47094  0.61335   4.029 0.000353 ***
GÉNERO         -0.24754  0.22988  -1.077 0.290134
CABEZAFMLIAi    1.90249  0.51977   3.660 0.000962 ***
EDAD            0.06402  0.03083   2.077 0.046478 *
SITUACIÓNHOGAR{temporal por razones de estudio} -0.72579  0.28402  -2.555 0.015917 *
NÚMERO.PERSOS.HOGAR 0.13285  0.07919   1.678 0.103798
NUMERO.DE.PERSOS.A.CARGO -0.32118  0.20359  -1.578 0.125151
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 52
Degrees of Freedom for the fit: 22
Residual Deg. of Freedom: 30
                        at cycle: 20

Global Deviance: 78.90009
AIC: 122.9001
SBC: 165.8275
*****
```


Figura 5-77: Diagnóstico de la distribución modelo Comunicación Escrita Factor Demográfico.

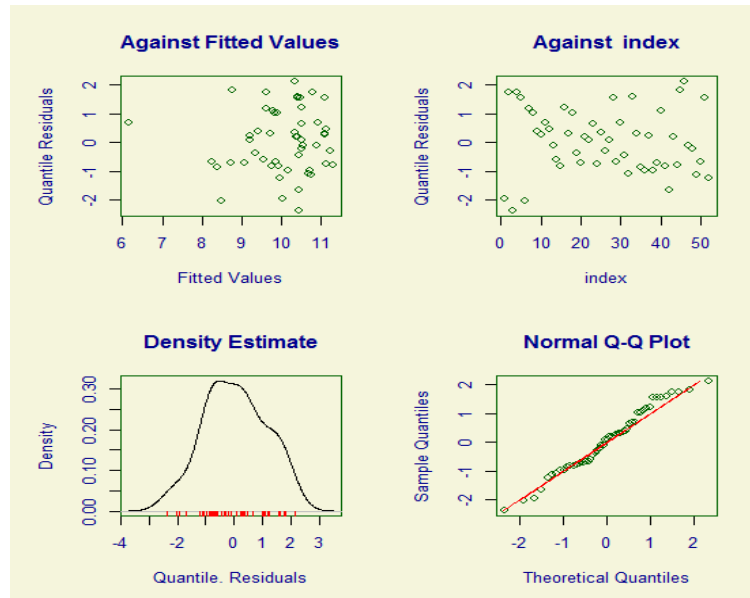
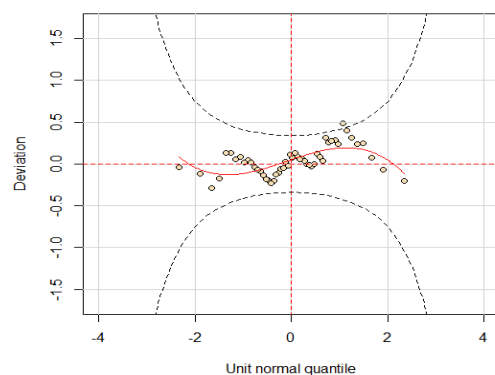


Figura 5-78: Prueba ajuste modelo Comunicación Escrita Factor Demográfico.

```
*****
Summary of the Quantile Residuals
      mean   = 0.03936409
      variance = 1.182335
      coef. of skewness = -0.0204776
      coef. of kurtosis = 2.256946
Filliben correlation coefficient = 0.9905273
*****
```

En las figuras 5-77 y 5-78 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-79: Ajuste de la distribución modelo Comunicación Escrita Factor Demográfico.**Figura 5-80:** Valores atípicos Número de personas a cargo Comunicación Escrita.

```
> Q.stats(w,xvar=Dataset$NUMERO.DE.PERSOS.A.CARGO ,n.inter=4)
```

	Z1	Z2	Z3	Z4	AgostinoK2	N
-0.5 to 0.5	0.44745837	-0.004321188	0.8793905	0.5384850	1.0632936	55
0.5 to 1.5	-0.39759817	-1.259777724	0.3636912	-0.7097005	0.6359461	21
1.5 to 5.5	0.04593877	1.447513506	-1.3653237	0.2879685	1.9470348	18
TOTAL Q stats	0.36041367	3.682353935	2.7697078	0.8765667	3.6462746	94
df for Q stats	1.00000000	2.00000000	2.0000000	3.0000000	5.0000000	0
p-val for Q stats	0.54827658	0.158630613	0.2503604	0.8310788	0.6013790	0

Figura 5-81: Valores atípicos Edad Comunicación Escrita.

```
> Q.stats(w1,xvar=Dataset$EDA,n.inter=4)
```

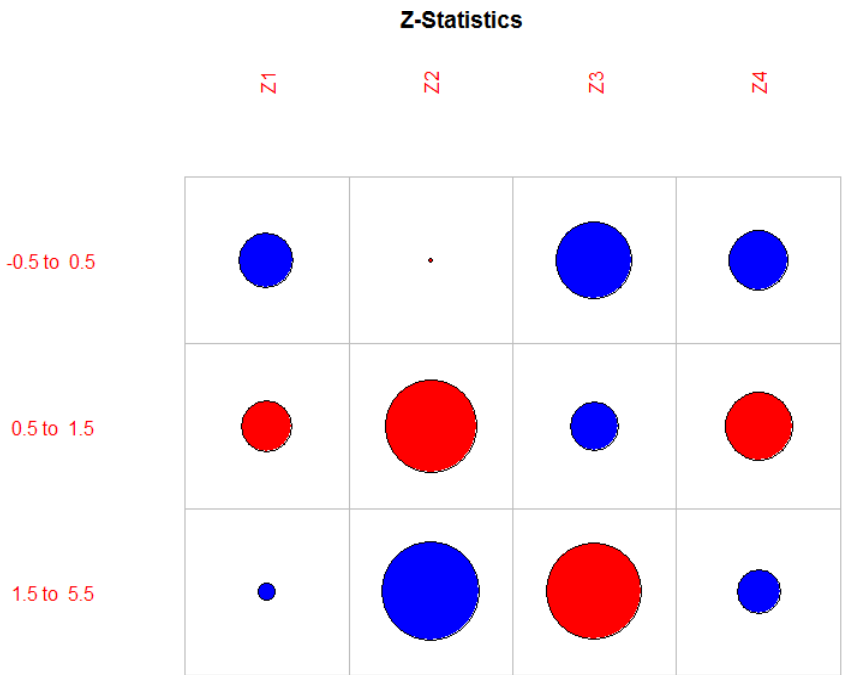
	Z1	Z2	Z3	Z4	AgostinoK2	N
19.5 to 24.5	0.7186429	0.9268229	-1.56509323	1.989918931	6.40929418	33
24.5 to 26.5	-0.9906376	-1.4149518	-1.64539937	-0.006579485	2.70738236	22
26.5 to 29.5	0.3923826	0.1338622	-2.27556251	1.774568972	8.32727975	18
29.5 to 51.5	-0.5067847	-1.3635553	0.27213514	-0.095023911	0.08308708	23
TOTAL Q stats	1.9086053	4.7382913	10.40909815	7.117945223	17.52704337	96
df for Q stats	2.0000000	3.0000000	4.0000000	4.00000000	8.00000000	0
p-val for Q stats	0.3850806	0.1919949	0.03407244	0.129785094	0.02506570	0

Figura 5-82: Valores atípicos Número personas hogar Comunicación Escrita.

```
> Q.stats(w2,xvar=Dataset$NÚMERO.PERSOS.HOGAR ,n.inter=4)
```

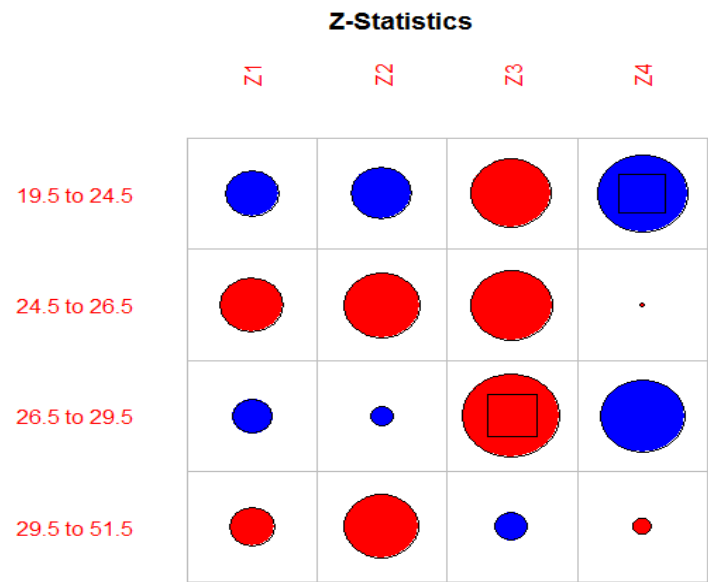
	Z1	Z2	Z3	Z4	AgostinoK2	N
0.5 to 3.5	1.62401928	-1.0292171	-0.311772454	-0.4509478	0.3005560	35
3.5 to 4.5	-1.62014092	-1.1408308	0.298728566	-1.3364906	1.8754458	22
4.5 to 6.5	-0.50367661	1.6170138	-0.005465077	0.6520216	0.4251621	29
6.5 to 9.5	0.93215449	-0.4583898	0.151310261	-1.3001823	1.7133689	8
TOTAL Q stats	6.38489733	5.1856375	0.209365481	4.1051673	4.3145328	94
df for Q stats	2.00000000	3.00000000	3.000000000	4.00000000	7.00000000	0
p-val for Q stats	0.04107118	0.1586977	0.976063414	0.3919601	0.7429200	0

Figura 5-83: Prueba de valores atípicos número de personas a cargo Comunicación Escrita.



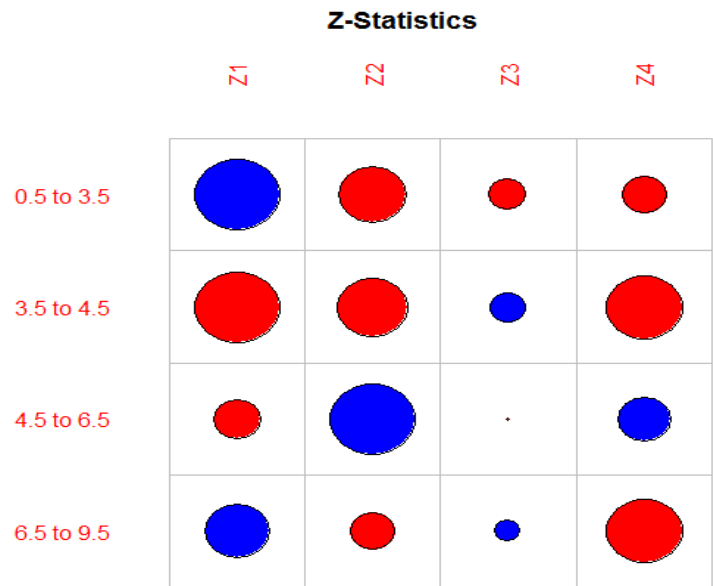
La variable número de personas a cargo no presenta datos atípicos.

Figura 5-84: Prueba de valores Atípicos Edad Comunicación Escrita.



La variable edad presenta datos atípicos, indicando una baja edad (-2.27).

Figura 5-85: Prueba de valores atípicos Número de personas en el hogar Comunicación Escrita.



La variable número de personas en el hogar no presenta datos atípicos.

A partir de la validación del modelo óptimo se puede decir que, al comparar dos estudiantes donde el segundo difiere del primero en tan sólo la situación sentimental se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes casados, el puntaje promedio de Comunicación Escrita disminuye 1.97 puntos respecto a los estudiantes en unión libre. Con un nivel de significancia inferior al 5 %, que aquellos estudiantes solteros, el puntaje promedio de Comunicación Escrita disminuye 2.62 puntos respecto a los estudiantes en unión libre.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el hecho de ser cabeza de familia, se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes que son cabeza de familia, el puntaje promedio de Comunicación Escrita disminuye en 1.89 puntos respecto a los estudiantes que no lo son.

Al comparar dos estudiantes donde el segundo difiere del primero en tan solo un año de su edad, se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que presenta mayor edad, obtienen en promedio una disminución de 0.14 puntos en el puntaje de Comunicación Escrita.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo una persona de más en el hogar, se tiene que, con un nivel de significancia del 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que presentan un mayor número de personas en el hogar, obtienen en promedio una disminución de 0.20 puntos en el puntaje de Comunicación Escrita.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo una persona de más a cargo, se tiene que, con un nivel de significancia inferior al 5 %, que aquellos estudiantes de Licenciatura en Matemáticas y Estadística que presentan un mayor número de personas en el hogar, obtienen en promedio un aumento de 0.38 puntos en el puntaje de Comunicación Escrita. También se determina que la variabilidad del puntaje de los estudiantes Licenciatura en Matemáticas y Estadística en Comunicación Escrita se ve explicado por el número de persona a cargo con un nivel de significancia inferior al 5 %.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo la situación civil se tiene que, con un nivel de significancia inferior al 5 %, que la explicación

del sesgo en el puntaje de Comunicación Escrita se da por estar casados respecto a los estudiantes en unión libre. Con un nivel de significancia inferior al 5 %, que la explicación del sesgo en el puntaje de Comunicación Escrita se da por estar solteros respecto a los estudiantes en unión libre.

Al comparar dos estudiantes donde el segundo difiere del primero en tan sólo el hecho de ser cabeza de familia se tiene que, con un nivel de significancia inferior al 5 %, que la explicación del sesgo en el puntaje de Comunicación Escrita se da por los que son cabeza de familia respecto a los estudiantes que no.

Al comparar dos estudiantes donde el segundo difiere del primero en tan solo un año de su edad, se tiene que, con un nivel de significancia del 5 %, que la explicación del sesgo en el puntaje de Comunicación Escrita se da por la edad de los estudiantes.

Al comparar dos estudiantes donde el segundo difiere del primero en tan solo su situación del hogar, se tiene que, con un nivel de significancia del 5 %, que la explicación del sesgo en el puntaje de Comunicación Escrita se da por los estudiantes con hogar temporal por razones de estudio respecto a los estudiantes con hogar habitual o permanente.

Modelo Factor Socioeconómico

Las Figuras 5-86 y 5-87 presentan el modelo óptimo correspondiente al Factor Socioeconómico.

Figura 5-86: Modelo óptimo Comunicación Escrita Factor Socioeconómico.

```
*****
```

```
Family: c("SN2", "skew normal type 2")
```

```
Call: gamlss(formula = COMUNICA_ESCRITA_PUNT ~ OCUP_MADRE1 + OCUP_PADRE1 + Recursos_forma.de.pago + ESTRATORECIBOENERGIA + VALR_MATRICULA_AÑO.ANTERIOR +  
CELULAR + INTERNET + TV + TELEFONIA + LAVADORA + FAMILIAR_MENSUAL + NÚMERODORMITORIOS + HORAS_TRABAJO, sigma.formula = ~1, nu.formula = ~1,  
family = SN2, data = na.omit(Dataset), trace = FALSE)
```

```
Fitting method: RS()
```

5.3 SELECCIÓN, EVALUACIÓN E INTERPRETACIÓN DEL MODELO ÓPTIMO 119

Continuación Modelo óptimo Comunicación Escrita Factor Socioeconómico.

Figura 5-87: Modelo óptimo Comunicación Escrita Factor Socioeconómico.

```
-----
Mu link function: identity
Mu Coefficients:

Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.97380 12.48914 -0.238 0.813
OCUP_MADRE1Obrero -7.01150 6.38475 -1.098 0.281
OCUP_MADRE1Ocupación no remunerado 0.06951 2.34522 0.030 0.977
OCUP_PADRE1No trabajador remunerado 0.03201 2.73113 0.012 0.991
OCUP_PADRE1Obrero -0.47996 2.55333 -0.188 0.852
OCUP_PADRE1Ocupación no remunerado 4.02682 5.14554 0.783 0.440
Recursos_forma.de.pagos1 2.92452 2.58497 1.131 0.267
ESTRATORRECIBOENERGIAESTR2 1.80827 4.26584 0.424 0.675
ESTRATORRECIBOENERGIAESTR3 0.17201 4.76865 0.036 0.971
ESTRATORRECIBOENERGIAESTR4 -3.16566 6.44865 -0.491 0.627
VALR_MATRICULA_AÑO.ANTERIOR(Entre 3 millones y 5 millones\r\r\n) -1.29615 9.26993 -0.140 0.890
VALR_MATRICULA_AÑO.ANTERIOR(Entre 500 mil y menos de 1 millón) 1.12731 6.49678 0.174 0.863
VALR_MATRICULA_AÑO.ANTERIOR(Más de 5 millones\r\r\n) 229.98655 229.71287 1.001 0.325
VALR_MATRICULA_AÑO.ANTERIOR(Menos de 500 mil\r\r\n) 3.11205 6.40682 0.486 0.631
CELULARsi 2.12363 4.58612 0.463 0.647
INTERNETsi 1.57120 2.27559 0.690 0.495
TVsi -0.03541 3.48530 -0.010 0.992
TELEFONIASi -3.93183 2.74566 -1.432 0.163
LAVADORAsi 1.66397 3.51049 0.474 0.639
FMIAR_MENSUAL(3 y 4 menos sm) -1.16113 3.09275 -0.375 0.710
FMIAR_MENSUAL1sm 1.53644 2.89198 0.531 0.599
NÚMERO_DORMITORIOS 1.47488 1.26171 1.169 0.252
HORAS_TRABAJO 0.05368 0.23383 0.230 0.820

-----
Sigma link function: log
Sigma Coefficients:

Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.25603 0.03829 6.687 1.7e-08 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Nu link function: log
Nu Coefficients:

Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.16164 0.03758 -4.302 7.67e-05 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 52
Degrees of Freedom for the fit: 25
Residual Deg. of Freedom: 27
at cycle: 20

Global Deviance: 174.6929
AIC: 224.6929
SBC: 273.474
*****
```

Figura 5-88: Diagnóstico de la distribución modelo Comunicación Escrita Factor Socioeconómico.

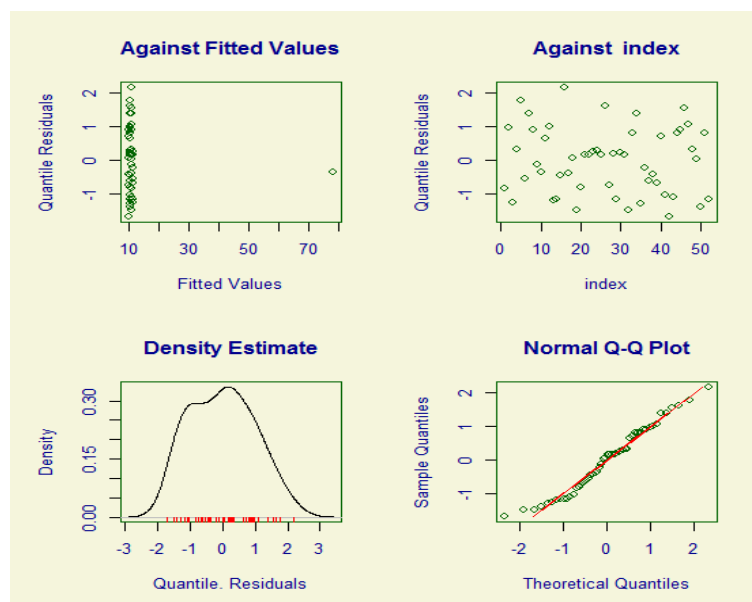
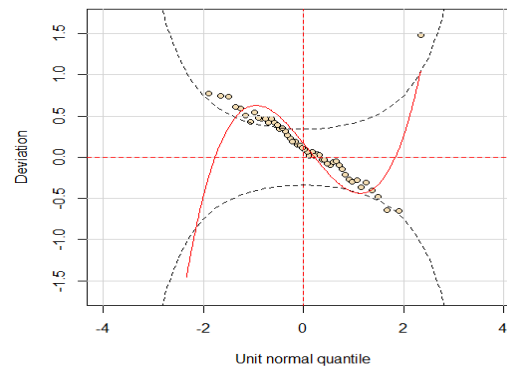


Figura 5-89: Prueba ajuste modelo Comunicación Escrita Factor Socioeconómico.

```
*****
Summary of the Quantile Residuals
      mean   = -0.01283238
      variance = 0.9639241
      coef. of skewness = 0.172392
      coef. of kurtosis = 2.046042
Filliben correlation coefficient = 0.9872774
*****
```

En las figuras 5-88 y 5-89 se muestran las pruebas gráficas que evalúan el modelo óptimo. Se observa que los residuales cuantiles aleatorizados se distribuyen como una normal, por lo tanto se puede concluir que estamos ante un modelo acertado.

Figura 5-90: Ajuste de la distribución modelo Comunicación Escrita Factor Socio-económico.



A partir de la validación del modelo óptimo se puede decir que, el factor socioeconómico no influye en el promedio, la variación y el sesgo del puntaje de Comunicación Escrita de los estudiantes de Licenciatura en Matemáticas y Estadística.

El factor Socioeconómico no es influyente en el puntaje de Comunicación Escrita que obtienen los estudiantes de Licenciatura en Matemáticas y Estadística en las pruebas Saber Pro. En el factor Demográfico el número de personas a cargo presenta un aumento en el promedio del puntaje mientras que el estado civil, la edad, ser cabeza de familia y el número de personas en el hogar, generan una disminución. En el factor Académico se ve una disminución en el promedio del puntaje, ocasionada por la educación de los padres (Primaria y secundaria).

5.3.5. Desempeño Comunicación Escrita

A continuación se presentarán los modelos correspondientes a los factores Académicos, Demográficos y Socioeconómicos del desempeño en Comunicación Escrita.

Modelo Factor Académico

Se encontró que el factor Académico no es influyente en el desempeño de Comunicación Escrita de los estudiantes de Licenciatura en Matemáticas y Estadística.

Modelo Factor Demográfico

La Tabla 5-4 presenta el modelo óptimo correspondiente al Factor Demográfico.

Tabla 5-4: Desempeño Comunicación Escrita Factor Demográfico

	Coef	Error	Wald	P-Valor	Odds	I.C. al 95 %
SEXO[T.M]	-0.686	0.386	-1.777	0.07876	0.5035	(0.2363235,1.073017)
Nivel1—Nivel2	-4.9076	1.0328	-4.7519	7.145e-06	0.0073	
Nivel2—Nivel3	-3.4832	0.5576	-6.2467	1.176e-08	0.0307	
Nivel3—Nivel4	-1.5265	0.3166	-4.8207	5.422e-06	1.2425	
Nivel4—Nivel5	-0.2255	0.2733	-0.8251	0.4113	0.79811	
Nivel5—Nivel6	2.8422	0.5307	5.3559	5.918e-07	17.1534	

Para la variable Sexo se observa que su razón de probabilidad y su intervalo de confianza al 95 % contiene al uno, lo que significa que el Sexo no tiene efecto sobre los niveles de desempeño en Comunicación Escrita.

Modelo Factor Socioeconómico

La Tabla 5-5 presenta el modelo óptimo correspondiente al Factor Socioeconómico.

Para la variable contar con servicio de Telefonía se observa que su razón de probabilidad y su intervalo de confianza al 95 % es menor que uno, lo que significa que la probabilidad acumulada, comenzando del Nivel de desempeño 1 hasta el 8, decrece en todos los valores que puede tomar Comunicación Escrita, así un estudiante que cuenta con servicio de Telefonía es menos probable que obtenga Niveles altos en el desempeño de Comunicación Escrita respecto a un estudiante sin este servicio.

Para la variable contar con servicio de Internet se observa que su razón de probabilidad y su intervalo de confianza al 95 % es mayor que uno, lo que significa que la probabilidad acumulada, comenzando del Nivel de desempeño 1 hasta el 8, aumenta en todos los

Tabla 5-5: Desempeño Comunicación Escrita Factor Socioeconómico

	Coef	Error	Wald	P-Valor	Odds	I.C. al 95 %
Telef[T.Si]	-1.274	0.571	-2.228	0.0281	0.2796	(0.09, 0.857)
Inter[T.Si]	0.992	0.488	2.031	0.0450	2.6982	(1.03, 7.03)
Trabaja[T.{Si,exprincia}]	-0.324	0.734	-0.442	0.6590	0.7225	(0.17, 3.04)
Trabaja[T.{Si,practica}]	3.290	2.044	1.609	0.1108	26.8589	(0.48, 1476)
TRABAJA[T.no]	-0.420	0.463	-0.907	0.3664	0.6565	(0.26, 1.62)
Ocupadre[T.no.T rab.Remu]	0.964	0.671	1.437	0.1538	2.624	(0.70, 9.77)
Ocupadre1[T.Obrero]	0.588	0.541	1.087	0.2796	1.8009	(0.62, 5.19)
Ocupadre1[T.Ocup.no Remu]	1.147	0.705	1.625	0.1073	3.1487	(0.78, 12.5)
Nivel1—Nivel2	-4.438	1.039	-4.269	4.649e-05	0.0118	
Nivel2—Nivel3	-2.982	0.571	-5.217	1.063e-06	0.0506	
Nivel3—Nivel4	-0.992	0.348	-2.845	0.0054	0.3705	
Nivel4—Nivel5	0.388	0.336	1.153	0.2514	1.4749	
Nivel5—Nivel6	3.644	0.610	5.964	4.17e-08	38.2445	

valores que puede tomar Comunicación Escrita, así un estudiante que cuenta con servicio de Internet es más probable que obtenga Niveles altos en el desempeño de Comunicación Escrita respecto a un estudiante que no cuente con este servicio.

5.3.6. Desempeño Inglés

A continuación se presentarán los modelos correspondientes a los factores Académicos, Demográficos y Socioeconómicos del desempeño en Inglés.

Modelo Factor Académico

La Tabla 5-6 presenta el modelo óptimo correspondiente al Factor Académico.

Para las variables Tipo de bachillerato (Normalista superior y Técnico) y Educación del padre (Primaria y Secundaria), se observa que sus intervalos de confianza al 95 % contiene al uno, lo que significa que estas variables no tienen efecto sobre el desempeño en Inglés.

Tabla 5-6: Desempeño Inglés Factor Académico

	Coef	Error	Wald	P-Valor	Odds	I.C. al 95 %
Tipbachill[Tecnico]	0.9809	1.1461	0.855	0.394	2.66	(0.282, 25.20)
Tipbachill[Normsupe]	0.6634	0.3965	1.673	0.097	1.94	(0.892,4.223)
Edupadre1[T.Primaria]	-0.4296	1.6161	-0.265	0.790	0.65	(0.027, 15.45)
Edupadre1[T.Secundaria]	-0.3933	1.6235	-0.242	0.809	0.67	(0.028, 16.26)
A—A1	-0.9987	1.6114	-0.619	0.536	0.36	
A1—A2	0.7287	1.6171	0.4506	0.653	2.07	
A2—B1	1.6621	1.6169	1.0280	0.306	5.27	

Modelo Factor Demográfico

La Tabla 5-7 presenta el modelo óptimo correspondiente al Factor Demográfico.

Tabla 5-7: Desempeño Inglés Factor Demográfico

	Coef	Error	Wald	P-Valor	Odss	I.C al 95 %
Núm.Persos.Hogar	-0.1724	0.117	-1.47	0.144	0.841	(0.668, 1.059)
Edad	-0.1278	0.047	-2.70	0.008	0.880	(0.802, 0.9655)
A—A1	-5.2319	1.508	-3.46	0.0008	0.005	
A1—A2	-3.4220	1.450	-2.35	0.020	0.032	
A2—B1	-2.4624	1.443	-1.70	0.092	0.085	

Para la variable Edad se observa que su razón de probabilidad y su intervalo de confianza al 95 % es menor que uno, lo que significa que la probabilidad acumulada, comenzando del nivel de desempeño A1 a B2, decrece en todos los valores que puede tomar Inglés, así un estudiante que tiene un año de más cumplido es menos probable que obtenga Niveles altos en el desempeño en Inglés.

Modelo Factor Socioeconómico

La Tabla 5-8 presenta el modelo óptimo correspondiente al Factor Académico.

Para las variable contar con servicio de televisión se observa que su intervalo de confianza

Tabla 5-8: Desempeño Inglés Factor Socioeconómico

	Coef	Error	Wald	P-Valor	Odds	I.C al 95 %
TV	0.985	0.5324	1.85	0.067	2.677	(0.943, 7.607)
A—A1	-0.539	0.3623	-1.48	0.142	0.583	
A1—A2	1.368	0.4106	3.33	0.001	3.927	
A2—B1	2.204	0.4858	4.53	1.71e-05	9.061	

al 95 % contiene al uno, lo que significa que contar con servicio de televisión no tiene efecto sobre los niveles de desempeño en Inglés.

6 DISCUSIÓN

A continuación las variables de los factores Académicos, Demográficos y Socioeconómico, que influyen en los puntajes de Razonamiento Cuantitativo, Inglés, Comunicación Escrita, Lectura Crítica y Desempeños en Inglés y Comunicación Escrita.

Las Tablas **6-1**, **6-2** y **6-3** muestran los factores que influyen a un nivel de significancia del 1 %, 5 % y 10 % para cada componente, siendo , R.C, Razonamiento Cuantitativo, I, Inglés, L.C, Lectura Crítica, C.E, Comunicación Escrita, D.I, Desempeño en Inglés, D.C.E, Desempeño Comunicación Escrita, “—”, indica disminución del puntaje por la variable, “+”, indica aumento del puntaje por la variable, los espacios en blanco indican que puntajes y desempeños no presentan incidencia por la variable, “**”, indica que decrece en todos los valores del desempeño y “*”, indica que aumenta en todos los valores del desempeño.

Tabla 6-1: Variables influyentes-Factor Académico.

Variables Factor Académico	R. C.	I	L. C	C. E	D. E.	D. I.
Educ.madre.Primaria			—	—		
Educ.padre.Primaria	—	—		—		
Educ.padre.Secundaria	—	—		—		
Educ.padre- Primaria				—		

De acuerdo a la Tabla **6-1**, se observa que la educación de los padres (Primaria y Secundaria) influye de manera negativa en el puntaje de cada componente, mientras que en los desempeños no se presentó incidencia por estas variables.

De acuerdo a la Tabla **6-2**, se observa que pertenecer al sexo masculino, ser cabeza de familia, y tener: Mayor edad, una situación del hogar temporal, un número mayor de

Tabla 6-2: Variables influyentes-Factor Demográfico.

Variables Factor Demográfico	R. C	I	L. C	C. E	D. C. E	D.I
Estado civil (Casados)		+		–		
Estado civil (Solteros)				–		
Sexo (M)	–	–				
Cabeza de familia (Si)	–		–	–		
Edad (Mayor)	–	+		–		**
Situación Hogar (Temporal)	–	+				
Número personas hogar (Mayor)	–	–	–	–		
Número de personas a cargo (Mayor)	–		+	+		

personas a cargo y en el hogar, aumenta el puntaje de Razonamiento Cuantitativo.

Para el puntaje de Inglés, se observa que estar casado, tener mayor edad y una situación del hogar temporal, aumenta el puntaje, mientras que pertenecer al sexo masculino y tener un número mayor de personas en el hogar lo disminuye. El desempeño en inglés decrece en todos los niveles al tener mayor edad.

Respecto a Lectura Crítica, se encontró que ser cabeza de familia y tener una mayor edad disminuye el puntaje, en cambio tener un número mayor de personas a cargo lo aumenta.

El ser cabeza de familia, casado o soltero con mayor edad, y tener un número mayor de personas en el hogar, disminuye el puntaje de Comunicación Escrita, mientras que tener un número mayor de personas a cargo lo aumenta.

En la Tabla **6-3**, se observa que las variables del Factor Socioeconómico no influyen en los puntajes de Razonamiento Cuantitativo, Comunicación Escrita y el Desempeño en Inglés.

Respecto al puntaje de Inglés, se observa que pertenecer al estrato dos, donde la ocupación de la madre es obrero, y tener: Celular, un mayor número de dormitorios y horas de trabajo, lavadora, trabajo y un valor de matrícula menor a \$ 500 mil y otro de tres a cinco millones de pesos, disminuye el puntaje.

Tabla 6-3: Variables influyentes-Factor Socioeconómico.

Variables Factor Socioeconómico	R. C	I	L. C	C. E	D. C. E	D.I
Estrato 2 Recibo energía		–				
Estrato 4 Recibo energía		+				
Valor anual matrícula. Menor 500 mil		–				
Valor anual matrícula. Entre 3 y 5 millones		–				
Recursos de pago						
Ocupación madre (Obrero)		–				
Ocupación padre						
Celular (Si)		–				
Internet (Si)					*	
TV (Si)		+	+			
Teléfono fijo (Si)		+			**	
Computador (Si)		+				
Lavadora (Si)		–				
Salario mensual. 4– 5 menos smmlv			+			
Número de dormitorios		–				
Horas de trabajo		–				
Trabaja (Para experienc.) (Si)		–				
Trabaja (Practica (Si)						

Tener televisión y un salario mensual familiar entre 4 y 5 S.M.V, aumenta el puntaje de Lectura Crítica.

Contar con servicio de internet aumenta el Desempeño en Comunicación Escrita, mientras que tener teléfono fijo lo disminuye.

Acorde con los antecedentes y los resultados obtenidos en esta investigación, se tiene:

De acuerdo con Castellanos *et al* (2014), existe una diferencia entre los puntajes obtenidos en Inglés por los estudiantes de diferentes estratos, como también una asociación con el desempeño en Comunicación Escrita. Además se encontró una diferencia en los promedios respecto al sexo en Razonamiento Cuantitativo, obteniendo puntajes más

altos las mujeres en este componente, de manera opuesta como lo menciona Castellanos *et al* (2014).

Al comparar resultados con Manrique, J. y Carreño, C. (2014), el sexo no afecta el desempeño en Comunicación Escrita, es decir no importa ser hombre o mujer para obtener niveles alto o bajos, ya que no es una variable influyente. Para el caso de internet, resulta tener una influencia positiva en el desempeño de Comunicación Escrita como concluyeron Manrique, J. y Carreño, C. (2014). De manera distinta con Manrique, J. y Carreño, C. (2014), la educación del padre (Primaria y Secundaria), si presento incidencia en el promedio del Desempeño en Comunicación escrita y en el puntaje de Inglés, evidenciado por el análisis bivariado, el modelamiento estadístico (GAMLSS) y la regresión logística ordinal (RLO) aplicada en el estudio.

Se recomienda una incorporación de Inglés con en el plan de estudios de tal manera que influya en el promedio acumulado, esto sería lo adecuado para fortalecer el desempeño en esta área y así lograr mejores resultados en este componente de la pruebas Saber Pro. Por otro lado, el internet es una herramienta relevante para el desarrollo general de las temáticas, contar con un apto funcionamiento de este en la Universidad, sería de gran ayuda para el desempeño de los estudiantes de Licenciatura en Matemáticas y Estadística.

7 CONCLUSIONES

En la población objeto de análisis se encontró que tienen una edad promedio de 27 años, en su mayoría son hombres, cuyos padres tienen mayoritariamente grado escolar de primaria completa, con una ocupación no remunerada en el caso de las madres y los padres con ocupación de administrador, siendo solteros, con un pago de matrícula menor a 500 mil pesos, pertenecientes a un tipo de bachillerato normalista, con situación del hogar permanente, donde el promedio de personas en el hogar es de cuatro, caracterizados por ser cabeza de familia, con una persona en promedio a cargo, ubicados en estrato dos en el recibo de energía y sisben, los cuales cuentan con celular, telefono fijo y lavadora, de manera opuesta no cuentan con internet ni televisión. El ingreso familiar esta entre dos y tres salarios mínimos legales vigentes, tienen en promedio tres dormitorios en el hogar, se desempeñan laboralmente como ayudantes sin remuneración trabajando en un promedio de 16 horas por semana, caracterizados por repasar por cuenta propia para la prueba Saber Pro.

Respecto a Razonamiento Cuantitativo se encontró diferencia entre los promedios según el sexo, el internet y el pago por cuenta propia de la matrícula. El puntaje no presenta incidencia por las variables explicativas seleccionadas en el modelo óptimo para el factor de Socioeconómico. El promedio y el sesgo se ven influenciados por las variables del factor Académico como educación del padre (Primaria y Secundaria), también por variables factor Demográfico como sexo, edad, número de personas en el hogar, número de dormitorios, ser cabeza de familia y el tipo de situación del hogar, variables como sexo, ser cabeza de familia, edad, número de personas a cargo y en el hogar, influyen en la variabilidad del puntaje. El sexo y el estado civil (Soltero) afectan el sesgo de Razonamiento Cuantitativo.

Se observó una diferencia entre los promedios de Inglés según la educación de la madre, el pago de matrícula por cuenta propia y de los padres y el nivel del sisben. El puntaje de Inglés no se ve influenciado en su variabilidad por el factor académico y Socioeconómico. El factor demográfico presenta alteraciones para el promedio y la variabilidad, variables como valor de la matrícula del semestre anterior, ocupación de la madre, estrato de recibo de energía, celular, telefonía, televisión, lavadora, número de dormitorios, horas de trabajo y trabaja. Para el factor Académico las variables como educación del padre (Primaria y Secundaria) genera cambios en el promedio del puntaje.

Se encontró una diferencia entre los promedios para Lectura crítica según la educación de la madre, ocupación del padre y pago de matrícula por cuenta propia. La variabilidad y el sesgo del puntaje de Lectura Crítica no son explicados por los factores: Académico, Demográfico y Socioeconómico, causado por problemas de convergencia en un modelo complejo. Variables como la educación de la madre, ser cabeza de familia, número de personas a cargo y en el hogar, servicio de televisión y tener un ingreso familiar mensual entre cuatro y cinco smmlv., inciden en el promedio de Lectura Crítica.

Se encontró una diferencia de medias de acuerdo al porcentaje de los créditos cursados en Comunicación Escrita. El promedio del puntaje en Comunicación Escrita, se ve influenciado por las variables educación de la madre (Primaria) y del padre (Primaria y Secundaria) del factor Académico, la educación de la madre presenta cambios en la asimetría del puntaje. El factor Socioeconómico no afecta el promedio, la varianza y el sesgo. El factor Demográfico con excepción de la variable situación del hogar, genera cambios en el promedio del puntaje.

Al relacionar los puntajes obtenidos en los módulos del componente genérico y algunas características de los estudiantes se encontró asociación entre el Desempeño de Comunicación Escrita y la educación de la madre, la educación del padre, el nivel de sisben, el porcentaje créditos cursados cuando presentó la prueba y el haber tomado un refuerzo en competencias genéricas. El sexo no afecta los niveles del Desempeño en Comunicación Escrita, por el contrario tener servicio de telefonía disminuye la probabilidad de obtener valores altos en el desempeño, en el caso de contar con servicio de internet aumenta la probabilidad de alcanzar niveles altos Comunicación Escrita.

Se encontró asociación entre el desempeño en Inglés y la educación de la madre, el haber tomado refuerzo en competencias genéricas y el pago matrícula el pago de la matrícula por cuenta propia. El tipo de bachillerato (Normalista superior y Técnico), la educación del padre (Primaria y Secundaria) y contar con servicio de televisión, no tienen incidencia en el Desempeño en Inglés. La edad influye negativamente, ya que al tener un año de más cumplido es menos probable que obtenga niveles altos en Inglés.

REFERENCIAS BIBLIOGRÁFICAS

- Agresti, A. y Kateri, M. (2011), Categorical data analysis, *in* ‘International encyclopedia of statistical science’, Springer, pp. 206–208.
- Báez Mancera, M. E. (2014), ‘Diseño de un modelo para evaluar el aprendizaje de los estudiantes de las instituciones de educación superior IES. caso de estudio ingeniería industrial universidad católica de colombia’.
- Castellanos, W. A. O., Perdomo, A. D. M. y Flórez, T. E. R. (2014), ‘Factores asociados al rendimiento en las pruebas saber pro en estudiantes de ingeniería civil en universidades colombianas’, *INNOVACIENCIA* **2**(1), 22–29.
- Cayuela, L. (2009), ‘Modelos lineales generalizados GLM’, *Materiales de un curso del R del IREC*.
- Congreso de Colombia (1992), *Ley 30*.
- Enrique, L. A. (2011), ‘Análisis comparativo de los factores académicos y sociales incidentes en el comportamiento del resultado de los exámenes de calidad para la educación superior 2005-2006 presentadas por los estudiantes de economía de la universidad de la salle’.
- Gonzalez, J. J. G. (2017, p 1- 31.), ‘Aplicación de un modelo de regresión logística con respuesta politómica ordinal en el análisis del desempeño académico en matemáticas’.
- ICFES (2009), ‘Guías orientaciones para el examen de estado de calidad de la educación superior saber pro ECAES Bogotá, DC’, ICFES .
- ICFES (2004 2009), ‘Guías de orientación, Bogotá D.C.’, ICFES.
- ICFES (2006- 2015), ‘Bases de datos, Bogotá D.C.’, ICFES.

- ICFES (2010), 'Guías orientaciones para el examen de estado de calidad de la educación superior saber pro ECAES medicina. Bogotá D.C.', ICFES .
- ICFES (2011a), 'Guías de orientaciones para el examen de estado de calidad de la educación superior SABER PRO ECAES', Guías, ICFES.
- ICFES (2011b), 'Orientaciones para el examen de estado de calidad de la educación superior saber pro (ecaes). Bogotá D.C.', ICFES.
- ICFES (2012-1 a 2013-2), 'Diccionario de variables saber pro pruebas genéricas periodos, (universitarios), Bogotá D.C.', ICFES.
- ICFES (n.d.a), 'Diccionario de datos saber pro ECAES. Bogotá D.C.', ICFES.
- ICFES (n.d.b), 'Diccionario de variables saber pro pruebas genéricas periodos 2012-1 a 2013-2 (universitarios)', ICFES. Bogotá D.C.
- ICFES (n.d.c), 'Programas por año Saber Pro, Bogotá D.C.', ICFES.
- ICFES (n.d.d), 'www.icfes.gov.co', ICFES.
- López-González, E. y Ruiz-Soler, M. (2011), 'Análisis de datos con el modelo lineal generalizado. una aplicación con r', *Revista Española de Pedagogía* pp. 59–80.
- Manrique, I. J. A. y Carreño, C. A. A. (2014), 'Influencia de los padres en el rendimiento académico de los hijos: una aproximación econométrica en el contexto de la educación media colombiana', *Revista de Educación y Desarrollo Social* 8(2), 184.
- McCullagh, P. (1980), 'Regression models for ordinal data', *Journal of the royal statistical society. Series B (Methodological)* pp. 109–142.
- MEN (2001a), *Decreto 1716*, Colombia.
- MEN (2001b), *Decreto 2233*, Colombia.
- MEN (2002), *Decreto 1373*, Colombia.
- MEN (2008), 'Análisis de determinantes de la deserción en la educación', MEN. Bogotá, Colombia.
- MEN (2009a), *Decreto 3963*, Colombia.
- MEN (2009b), *Ley 1324*, Colombia.
- Mesa, H. D. V. (2009), 'De los ecaes, la autonomía universitaria y el derecho', *Diálogos de Derecho y Política* .

- OCDE (2016), 'Revisión de políticas nacionales de educación: Educación en Colombia', OCDE.
- Peña, M. J. P. (2015-2016), 'Modelos aditivos generalizados para localización, escala e forma GAMLSS'.
- Perlaza Vásquez, A., Rodríguez Cordero, C. Y. y Uribe Gómez, P. P. (2009), Fortalezas y debilidades de estudiantes de último año de licenciaturas, frente a las competencias transversales evaluadas en el ECAES, Master's thesis, Facultad de Educación.
- Rodríguez, O., Casas, P. P. y Medina, Y. (2005), 'Análisis psicométrico de los exámenes de evaluación de la calidad de la educación superior ECAES en Colombia', *Avances en Medición* **3**, 153–172.
- Stasinopoulos, M., Rigby, B. y Akantziliotou, C. (2008), 'Instructions on how to use the gamlss package in R second edition'.
- Tolosa, F, R. H. . R. D. (2014, p 34,35), 'El papel de la evaluación de competencias en los procesos de formación docente: análisis de las pruebas saber pro para licenciados de psicología y pedagogía, licenciados en español y lenguas extranjeras en la universidad pedagógica nacional'.
- Tovar, D. A. (2013), 'La universidad colombiana en el período neoliberal (1990-2013)'.